National School of Public Health
NOVA UNIVERSITY LISBON

Variations in Medical Practices

Identification, Causes and Consequences

Doctoral Programme of Public Health

Specialization in Health Economics

Maria Inês Rodrigues dos Santos Joaquim

May, 2018

National School of Public Health
NOVA UNIVERSITY LISBON

Variations in Medical Practices

Identification, Causes and Consequences

This thesis is presented as part of the requirements for the Degree of Doctor of Public Health, under supervision of Prof. Céu Mateus and Prof. Carla Nunes

May, 2018

Tutorial Commission

Prof. Céu Mateus

Division of Health Research, Lancaster University

Prof. Carla Nunes

Escola Nacional de Saúde Pública, Universidade Nova de Lisboa

Prof. Luís Nobre Pereira

Escola de Gestão, Hotelaria e Turismo, Universidade do Algarve

Agradecimentos

Não há como não agradecer em primeiro lugar à Prof. Céu Mateus. A minha gratidão vai além da orientação desta tese de doutoramento, uma gratidão que nunca lhe expressei. O convite para trabalhar na Escola Nacional de Saúde Pública, em Novembro de 2010 surgiu num dos momentos mais infelizes pelos quais passei, com toda a carga emocional que representou. Com esta mudança, viria a perspectiva de trabalhar na área que mais me entusiasmava, a área da saúde, viria a perspectiva de realizar o doutoramento e viria um dos períodos de maior crescimento pessoal. A Prof. Céu Mateus foi a responsável pelo muito que aprendi sobre economia da saúde neste período, proporcionando-me o contacto com economistas da saúde de outras instituições internacionais no projecto ECHO, proporcionando a minha participação em diversas Conferências de Economia da Saúde Nacionais e Internacionais e dando-me o privilégio de aprender directamente com ela. Para além da economia da saúde, transmitiu-me os valores de um trabalho bem feito (seja ele em que área), a responsabilidade, o brio e a resiliência necessárias para vingar. Foi, sem dúvida, essencial para que esta tese fosse realizada. Orientou, desafiou, fez a tese acontecer. Quem eu sou hoje, devo em parte à Prof. Céu Mateus.

Em segundo lugar, não posso deixar de agradecer à minha orientadora Prof. Carla Nunes. Mesmo que a ordem dos acontecimentos não ma trouxesse como orientadora "oficial", a Prof. Carla Nunes era já minha orientadora "oficiosa". Fazendo parte da equipa portuguesa do projecto ECHO esteve sempre presente no desenvolvimento dos trabalhos da tese. A Prof. Carla Nunes foi o motor para que esta tese chegasse ao fim, comprometendo-me, responsabilizando-me, mas acima de tudo, nunca desistindo de mim e dando sempre um pouco de si para compreender os dias menos bons. Desistir nunca foi opção mas a sua paciência infinita, a compreensão e o carinho demonstrados fizeram a diferença quando o que mais apetecia era fazer parar o processo.

This PhD thesis was integrated in the European Collaboration for Health Optimization (ECHO) project funded by the European Commission's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°242189. ECHO project was also the promoter of the international database that was used in this PhD thesis and without this database this PhD thesis would not be possible. To all ECHO colleagues, it was a privilege to meet you, to take part of the discussions and to learn with your experience: Anne-Marie Yazbeck, Carla Nunes, Céu Mateus, Enrique Bernal-Delgado, Jeni Bremner, Jörg Munck, Karen Bloor, Lau Thygesen, Nils Gutacker, Olivia Dix, Paul Giepmens, Ramón Launa-Garcés, Richard Cookson, Sandra Garcia-Armesto and Terkel Christiansen.

Não posso deixar de agradecer ao Prof. Luís Nobre Pereira que integrou a minha comissão tutorial (em conjunto com a Prof. Céu Mateus e a Prof. Carla Nunes) pelos comentários e sugestões.

Um agradecimento também ao Prof. Pedro Pita Barros pelos comentários detalhados ao trabalho "*The impact of hospital characteristics on the avoidable c-sections*".

Um estudante de doutoramento nunca está só e nunca não tem nada para fazer, a tese é uma presença constante. Amigos, colegas e familiares convivem não apenas com o estudante de doutoramento mas também com a sua tese e por isso, há várias pessoas que não tendo tido uma contribuição directa para a tese de doutoramento fizeram a diferença neste processo e a quem eu quero agradecer.

À minha grande mestre Mónica Inês que acreditou sempre que eu poderia ter um lugar na Economia da Saúde em Portugal, que me proporcionou muitas gargalhadas e boa disposição e nunca deixou de me motivar.

Aos meus colegas de doutoramento, Hugo Lopes e Vanessa Ribeiro e às minhas colegas da Escola Nacional de Saúde Pública, Joana Alves, Klára Dimitrovová, Cláudia Furtado, Patrícia Marques e Carolina Santos. São possivelmente as pessoas que melhor compreendem o processo, seja porque o estavam a viver em simultâneo, seja porque já o tinham vivido. A partilha de experiências e o seu apoio foi fundamental.

A todos os colegas da Medtronic que vendo-me a fazer serão sempre exprimiram uma palavra de incentivo e que sempre me motivaram a terminar o doutoramento, Inês Lucena, Joana Albuquerque, Margarida Noronha, Paulo Almeida e Susana Pereira. Um especial agradecimento ao Luís Pereira que foi não só um motivador e um conselheiro mas também criou uma competição saudável entre nós que me fez avançar nos trabalhos.

À minha família do coração por estarem lá quando é preciso, Ana Freitas, Clara Duarte, Cláudia Canedo, Filipe Abreu, Inês Almeida, Liliana Elsig, Rita Silva e Tânia Barquinha.

Devo também um agradecimento especial à minha colega e amiga Joana Alves. O que diga vai ser sempre pouco para expressar a gratidão que tenho para com ela. A Joana tornou-se mais do que minha colega de gabinete. Foi amiga, foi confidente, foi apoio moral e foi o meu mais prestável recurso para resposta a perguntas que me levariam bastante tempo a ser respondidas. Mas a Joana foi, e é, sobretudo, um grande exemplo para mim, tanto a nível pessoal como a nível profissional. Obrigada por tudo!

Aos meus pais, que me ensinaram o privilégio de estudar, e aos meus irmãos, agradeço toda a educação e conselhos que me deram ao longo da vida, a compreensão pelas ausências, o apoio e a paciência nos dias menos bons. São eles o meu maior exemplo de amor incondicional.

João Maria, Duarte, Anita e Manuel, dedico-vos esta tese para que se recordem sempre do valor do trabalho, do estudo e da concretização dos projectos a que se propõem. Os vossos sorrisos, abraços e beijos são os melhores do mundo!

Abstract

Background: Unwarranted variations in healthcare are thought to describe healthcare provision beyond what is clinically necessary and without additional clinical benefits raising concerns on quality, equity and efficiency of healthcare systems. This thesis aims to 1) identify geographical variations in Portugal and research on the potential of identification of geographical variations to optimize care; 2) understand how hospital characteristics affect provision of healthcare; 3) use stochastic frontier analysis to estimate innefficiencies resulting from non-optimal care and compare hospitals.

Methods: Firstly, nine healthcare activities performed in Portuguese National Health Service hospitals between 2002 and 2009 were analyzed according to area of residence of patients. Secondly, low-risk c-sections geographical variation and excess consumption was compared between five European countries. Thirdly, avoidable c-section rates are computed for Portuguese hospitals and hospital characteristics are studied to understand how they affect those rates. Fourthly, advantages and drawbacks of stochastic frontier analysis method for healthcare efficiency measurement are studied. Fifthly, stochastic frontier analysis is applied to healthcare activity of four European countries to compare hospital efficiency levels within and between countries.

Results: Variations in medical practice exist in Portugal but their magnitude and evolution varies with the procedure in analysis. Portugal's performance on international comparisons depend on the scope of the procedure. Availability of resources affect medical practice in a modest extent. SFA is a good analytical tool to compare hospital's efficiency levels. Efficiency levels of Portuguese hospitals are not homogenous even though inefficiencies can be attributed to random shocks out of hospital control.

Conclusions: The identification of variations in medical practice provides signals on where clinical harmonization is required while international comparisons provide benchmarking that flags improvement opportunities. Resources affect medical practice in a modest extent and so, policies on resource affection may result in modest results. Policies on medical incentives towards an objective may be more effective than common policies on resource reduction. More than looking towards an optimal threshold of care we shall guarantee that care is provided to who can benefit from it and look towards clinical outcomes optimization.

Keywords: variations in medical practice, small-area variations, efficiency, stochastic frontier analysis.

Resumo

Contexto: Identificar variações não desejadas da prática médica é identificar cuidados de saúde prestados além do que é clinicamente necessário e sem benefício clínico adicional, com implicações ao nível da qualidade, equidade e eficiência do sistema de saúde. Esta tese tem como objectivos: 1) identificar variações geográficas da prática médica em Portugal e potenciar essa identificação para optimização dos cuidados prestados; 2) compreender como é que as características dos hospitais afectam os cuidados prestados; 3) utilizar a análise de fronteira estocástica no cálculo da ineficiência gerada pela não-optimização dos cuidados e comparar hospitais.

Métodos: Primeiro foram analisados nove procedimentos realizados nos hospitais do Serviço Nacional de Saúde entre 2002 e 2009 tendo em consideração o local de residência dos doentes. De seguida, considerando apenas as cesarianas de baixo risco foi analisado para além das variações geográficas, a realização em excesso deste tipo de procedimentos e comparados os resultados de cinco países europeus. Seguiu-se uma análise às taxas de cesarianas evitáveis e de que forma as características dos hospitais se relacionam com estas. Por fim, foram estudadas as vantagens e limitações da análise de fronteira estocástica no cálculo da eficiência hospitalar e aplicou-se este método no cálculo da eficiência dos hospitais de quatro países europeus.

Resultados: Em Portugal, existem variações da prática médica embora a sua magnitude e evolução varie conforme o procedimento em análise. O desempenho de Portugal em comparação com outros países europeus também depende do procedimento e da própria definição do âmbito do procedimento. A disponibilidade de recursos afecta a prática médica de forma modesta. A análise de fronteira estocástica é uma boa ferramenta para estimar e comparar níveis de eficiência dos hospitais. Em Portugal, os níveis de eficiência não são homogéneos entre hospitais embora as ineficiências possam ser atribuídas a factores externos e aleatórios fora do controlo dos hospitais.

Conclusões: A identificação das variações da prática médica indica as áreas onde algum tipo de harmonização clínica é necessária e as comparações internacionais identificam áreas com potencial de melhoria. A disponibilidade de recursos afecta a prática médica de forma modesta pelo que políticas de saúde de redução de recursos poderão, também elas, ter resultados modestos. Políticas de saúde com incentivos dirigidos aos médicos poderão ser mais eficientes na homogeneização da prática médica. Mais do que definir um nível de actividade óptimo urge garantir que os cuidados de saúde são dirigidos a quem beneficia deles optimizando os resultados clínicos.

Palavras-chave: variações da prática médica, variações geográficas, eficiência, análise de fronteira estocástica.

Contents

List of Tables

List of Figures

List of Acronyms

ACHI	Australian Classification for Health Interventions
A-CS	Avoidable C-section
A-CSR	Avoidable C-section rate
AP-DRG	All-Patient Diagnostic Related Groups
APR-DRG	All-Patient Refined Diagnostic Related Groups
AR-DRG	Australian Refined Diagnostic Related Groups
ATLAS-VPM	Atlas de Variaciones en la Práctica Médica
CABG	Coronary Artery Bypass Grafting
C-Sections	Caesarean Section
CSR	C-section rate
CV	Coefficient of Variation
DEA	Data Envelopment Analysis
Dk NordDRG	Danish patient classification system
DRG	Diagnostic Related Groups
EB	Empirical Bayes
ECHO	European Collaboration for Health Optimization
EuroHOPE	European Healthcare Optimization, Performance and Efficiency
HRG	Healthcare Resource Groups
ICD-10	International Classification of Diseases, 10th version
ICD-10-CM	International Classification of Diseases, 10th version, Australian Modification
ICD-9-CM	International Classification of Diseases, 9th version, Clinical Modification
ICU	Intensive Care Unit
INE	Instituto Nacional de Estatística
NCSP	NOMESCO Classification of Surgical Procedures
NHS	National Health Services
NICE	National Institute for Health and Care Excellence
NOMESCO	Nordic Medico-Statistical Committee
NTSV	Nulliparous, Term, Singleton, Vertex
OECD	Organization for Economic Cooperation and Development

OPCS	Office of Population Censuses and Surveys
PID	Physician-Induced Demand
PROMs	Patient Reported Outcomes Measures
PTCA	Percutaneous Transluminal Coronary Angioplasty
SAV	Small Area Variations
SCV	Systematic Component of Variation
SFA	Stochastic Frontier Analysis
SID	Supply-Induced Demand
UK	United Kingdom
VR	Variation Ratio
W1	Work 1
W2	Work 2
W3	Work 3
W4	Work 4
W5	Work 5
WHO	World Health Organization

1. Introduction

Jack Worm, 63 years-old lives in Strawberry Fields in Fruitland. In another city of Fruitland, Plum Mountain, lives John Bug also 63 years-old. In a rainy day, after the Great Harvest Festival, the most important holiday in Fruitland, both are hospitalized for an acute ischemic stroke (AMI), Jack Worm in Hospital of Strawberries and John Bug in Saint Plum's Hospital. Although it was expected the same treatment for both Jack and John, treatments, length of stay and outcomes differed without any apparent reason apart from medical decision.

Differences in medical treatment were first noticed by Glover (1938) but it was in the 70s that this topic has started to boom (Glover, 1938; Wennberg, 2014). After the work of Wennberg & Gittelsohn (1973) which identified a great variation of tonsillectomies across the State of Vermont, a large number of other studies have found identic differences in several procedures (Corallo *et al.*, 2014; Srivastava *et al.*, 2014; Wennberg e Gittelsohn, 1973; Wennberg, 2014). For the last 40 years researchers have tried to explain the phenomena nominated unwarranted variations – variations that are not desired – in healthcare provision both geographically and hospital based creating a vast literature and a specific literature topic called Small Area Variations (SAV).

These unwarranted variations are thought to describe healthcare provision beyond what is clinically necessary – and without additional clinical benefits – raising concerns on the quality, equity and efficiency of healthcare systems (Peiró e Maynard, 2015; Srivastava *et al.*, 2014).

The studies evolved from the identification of variations in specific procedures to the identification of variations on overall healthcare services and the study of underlying differences in populations as well as the causes of SAV.

Reasoning behind SAV is not consensual because while some researchers state that geographical variations are a matter of population differences others find that even controlling for differences in population, much variation remains to be explained.

For those who state that variations are not a matter of population-based differences, a substantial research on physician's incentives to variations in clinical practice has been generated. This is called Supply-Induced Demand Theory and states that physicians are incentive-driven and sometimes broke their agency relationship with patients by taking decisions in their own best interest instead of deciding in the best interest of patients.

If, in fact, the underlying factors affecting geographic variations have less to do with populations' characteristics and more with supply factors such as clinical practice, then policies should focus not on the demand side but on the supply side. Understanding the causes for geographical variations is essential to design effective policies on the reduction of these variations.

While the assessment of geographical variations was first studied to understand the differences in healthcare services provision, these studies proved to be a tool to understand whether there is inappropriate care that is source of inefficiency and waste. Providing value for the money invested is essential when scarcity of resources and the need to become even more efficient are the focus of decision-makers.

SAV studies have revealed that some geographical areas are over or under treating their populations inducing welfare losses. Although both under and overuse induce welfare losses, the fast growth of the health expenditure of developed countries suggest that overuse is playing a greater role in resource wasting. In fact, in the last decades, spending in healthcare sector has been increasing without better results or improved quality in the services provided. Estimates suggest that about 30% of all healthcare spending derive from services that could have been avoided (Berwick e Hackbarth, 2012).

Geographical variations, if not justified by population characteristics, generate inefficiencies in healthcare institutions and healthcare systems as a whole (Pauly, 1980). Comparing similar geographies, institutions and healthcare systems provide some tools to address the efficiency or inefficiency compared to peers.

Although international comparisons on efficiency of healthcare systems have been developed since some decades on a macro level (Oxley e MacFarlen, 1994), the lack of comparable patient-level data has limited the development of international comparisons of efficiency of healthcare systems at a micro-level. Different approaches have been used trying to overcome these limitations, namely through research projects based on international datasets. As examples, two European projects are here referred: European Collaboration for Health Optimization (ECHO) and European Health Care Optimization, Performance and Efficiency (EuroHOPE). These two projects worked on the development of international patient-level databases with comparable data (ECHO, EuroHOPE; Peiró & Maynard, 2015; Sund & Häkkinen, 2016).

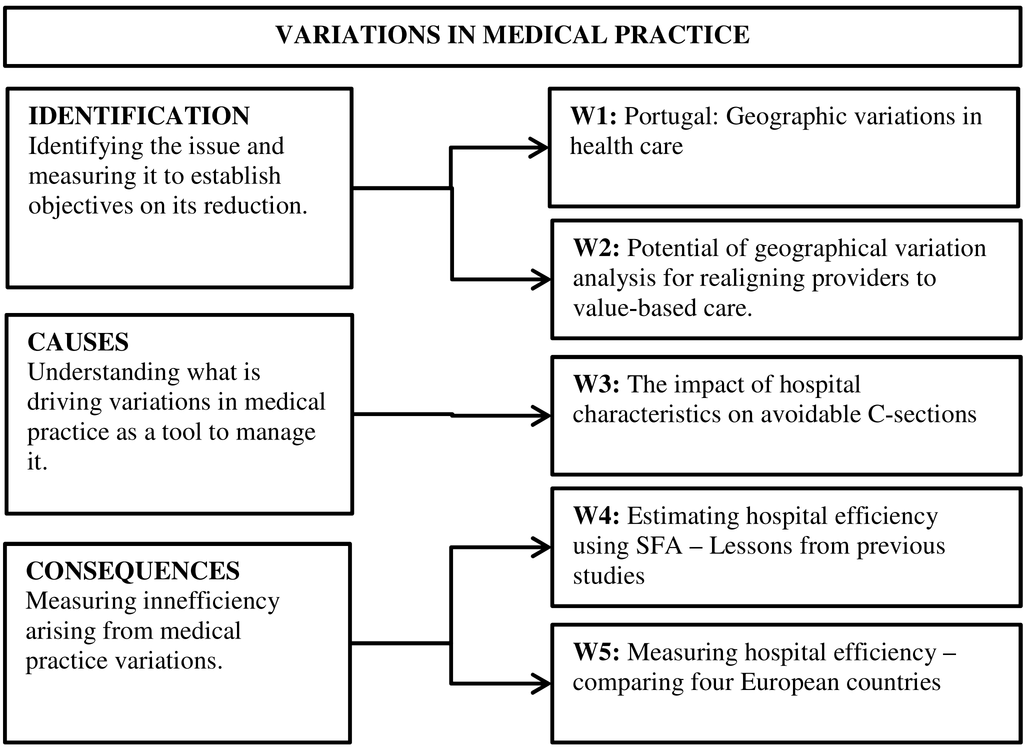
This thesis aims to fill in two gaps in the literature of geographical variations: 1) to estimate geographical variations in healthcare in Portugal; 2) to produce international comparisons on procedures rates and efficiency that were limited until now. To fulfill these objectives the international database developed by ECHO is used.

This thesis is structured in three sections to address variations in medical practice: identification, causes and consequences of this public health problem (Figure 1).

Identification framework recognizes the issue; whether there are variations or not and measures the magnitude of variations. Studying the causes helps to understand the reasoning behind the variations, what the incentives are and what is driving the variations. Finally, consequences of variations in medical practice are studied in the form of efficiency measures.

The study of these three elements resulted in five works (W) that provide an integrated view of the variations in medical practice (Figure 1). For the identification section there is a work that identifies geographical variations in Portugal (W1) and research on the potential of identification of geographical variations to optimize care (W2). A work on the impact of hospital characteristics on avoidable C-sections (W3) contributes to the question on the causes for geographic variation. Finally, two works on Stochastic Frontier Analysis (SFA) have been developed to estimate the consequences of geographic variations. A more theoretical work (W4) studies this methodology for efficiency estimation and then this methodology is applied to compare hospitals from 4 countries (W5).

Figure 1: Structure of the thesis



This PhD thesis is integrated in the ECHO project that was developed at the Escola Nacional de Saúde Pública. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°242189. Sole responsibility lies with the authors, and the European Commission is not responsible for any use that may be made of the information contained therein. Some work presented in this thesis is work of the author in collaboration with the other partners of the project. Nevertheless, it is original work developed by the author in her collaboration with the project mentioned above.

2. Background

This chapter is organized in two different sections: Small Area Variations (SAV) and Efficiency.

In the first section, an introduction to the history of SAV research is provided as well as methodologies and the main findings on the topic. Supply-induced demand is also explored and a section on different approaches to reduce SAV concludes the topic.

The second section defines efficiency and identifies its different approaches. The section continues with methodologies to estimate efficiency and to compare healthcare units and how outputs and inputs can be estimated.

2.1. Small Area Variations

Theory on SAV has been developing since the 70s with assessment of variations in resource input, utilization of services and expenditures between neighbor communities. The work of Glover (1938) on rates of tonsillectomy in England had already identified wide variations that were not justified by population or ill-defined characteristics but instead seemed to be based on medical opinions and practices. But it was the work of Wennberg & Gittelsohn (1973) that launched the basis for the development of SAV as research topic in the last forty years (Chandra, Cutler e Song, 2011; Glover, 1938; Peiró e Maynard, 2015; Srivastava *et al.*, 2014; Sundmacher e Busse, 2014; Wennberg e Gittelsohn, 1973).

In the words of Wennberg (2014), the uncovering of variations found in the State of Vermont led him to think that the problems of the healthcare systems were more deep than the barriers to the diffusion of technologies and innovations. In fact, there was much variation that could not be explained by illness, patient's preferences or medical science. Researchers named these variations – *unwarranted variations* (Wennberg e Fowler, 1977; Wennberg e Gittelsohn, 1973).

A contribution to the expansion of this topic was the creation of The Dartmouth Atlas of Health Care as a strategy to create awareness on the topics of geographical variation (Peiró e Maynard, 2015; Wennberg e Cooper, 1996; Wennberg, 2014). The works developed intended to characterize and measure the variations and to improve the understanding of the causes and consequences of variation in the delivery of healthcare. Additionally, works on the topic would expand policy research on how to address these variations, namely, through the development of theory of supply-induced demand (Peiró e Maynard, 2015; Wennberg, 2014).

The attention given to this topic in the United States shortly became noticed in other countries which understood the importance of the topic. During the 90s, countries like

Australia, New Zealand and Canada have also analyzed practice variations as well as European countries such as Spain, England and Germany which have created their own Atlases[1] during the last decade (OECD, 2014; Peiró e Maynard, 2015; Wennberg, 2014). The Organization for Economic Cooperation and Development (OECD) has also studied and compiled information on 13 countries in a recent report (OECD, 2014).

A common effort of several European countries through the project ECHO has given a step further by combining national hospital databases in order to move from a national perspective to a cross-country perspective enabling international comparisons of healthcare (Bernal-Delgado *et al.*, 2015; ECHO, 2014; Peiró e Maynard, 2015).

In Portugal, studies on this topic have been inexistent until the emergence of the two projects mentioned above in which Portugal was involved (Bernal-Delgado *et al.*, 2015; ECHO, 2014; Mateus *et al.*, 2014).

Variations in medical practice are not, in most of the cases, desirable. Variations usually express differences in clinical practice, successful increase of healthcare provision that may not be necessary or inequity in access to healthcare generating concerns on the equity and efficiency of healthcare systems (Cylus, Papanicolas e Smith, 2016; Srivastava *et al.*, 2014). In any case, variations may indicate that inappropriate care is being delivered (Cylus *et al.*, 2016; Folland, Goodman e Stano, 2013).

Healthcare systems evaluation – effectiveness, quality, efficiency and equity, though, is usually done through aggregate indicators which undermine the correct evaluation of public health policies and difficult the identification of areas or providers that may be generating this inappropriate care (Baicker e Chandra, 2004; Chandra *et al.*, 2011; Fisher *et al.*, 2003; Ridao-López *et al.*, 2012). Analysis of SAV disentangles these aggregate indicators and identifies individually the results of healthcare policies generating important information for policies evaluation that cannot be correctly assessed by analysis of aggregate indicators (Ridao-López *et al.*, 2012). If ineffective care is not correctly identified, determined and measured information is not correctly telling whether the investment is worthy (Ridao-López *et al.*, 2012).

These analyses are relevant in a context where healthcare systems are more and more pursuing value for money and making sure their investments are producing valuable outcomes for patients and populations.

---

[1] "Atlas de variaciones en la práctica médica en el Sistema Nacional de Salud, 2006 [The Atlas of Medical Practice Variation in the Spanish National Health System]"
"Medizinischen Versorgungatlas in Deutschland, 2011 [Atlas of Medical Care in Germany]"
"NHS Atlas of Variations in Healthcare, 2010"

If variations in healthcare are indicative of inappropriate care then these variations are a signal of poor quality patient care with implications on safety and public health (Peiró e Maynard, 2015).

From the point of view of public health, inappropriate care might mean that some populations are being exposed to excess healthcare without clear benefits and potential risks or, on the other hand, populations are being undertreated and lacking healthcare services they need. In any of the cases, losses are being incurred: populations are not receiving the healthcare they need or valuable resources are being wasted or could be better allocated (Cylus e Pearson, 2016; Sirovich *et al.*, 2006; Wennberg, 2002). Additionally, geography, only, should not be the determinant of healthcare utilization. Given the concern with the health of the community as a whole, it is of public health interest that all populations are served similarly – given the populations characteristics – at an optimal point.

From a health economics point of view, identifying regional variations in medical practice is to recognize that there are unwarranted variations and health spending that may not be translating into additional outcomes and are inducing welfare losses (Peiró e Maynard, 2015). These non-optimal practices are a source of waste and inefficiencies. The reduction of these unwarranted variations could potentially reduce healthcare spending by eliminating healthcare provision that is not necessary (without affecting patient's health), redirect these resources to beneficially treat other patients and improving the quality of care (Colla *et al.*, 2014; Folland *et al.*, 2013; Peiró e Maynard, 2015; Srivastava *et al.*, 2014).

Studies on SAV account for measures of variation across geographical units or providers but this research topic has tried to understand the reasoning behind these variations. Analyses focus on the geography of patients or the institutions they are treated in and look for reasons for the differences in practice, resources availability, reimbursement and financial incentives, factors that differ more than differences in patient population, differences in medical need or differences in patients' own socioeconomic conditions (Ridao-López *et al.*, 2012; Wennberg e Fowler, 1977; Wennberg e Gittelsohn, 1973). While some variation can be expected given case-mix and patient's preferences, SAV studies have the potential to address differences that depend solely on factors other than demand itself (Chandra *et al.*, 2011; Wennberg, Freeman e Culp, 1987; Wennberg e Gittelsohn, 1982).

Advantages of these studies focus mainly on the possibility of controlling for population's differences across regions which results in the analyses of differences in access only which may be the result of inefficiencies in the utilization of resources (Colla *et al.*, 2014; Ridao-López *et al.*, 2012). Moreover, the methods used in SAV studies clearly address variations that are systematic and variations that are random making it possible to study the systematic part of variation whether it is in utilization rates or healthcare spending (Ridao-López *et al.*, 2012).

SAV studies enable not only the measurement of the variations across regions but also the identification of these geographical areas, or providers, that consistently and systematically provide care above the expected (Bernal-Delgado, García-Armesto e Peiró, 2014; Colla *et al.*, 2014; Ridao-López *et al.*, 2012). This is especially relevant when talking about procedures that do not bring additional benefit and may be inputting additional costs on the healthcare systems systematically. Corrective policies can be designed when a problem is correctly identified and this information is of major importance to policies' evaluation (Bernal-Delgado *et al.*, 2014; Colla *et al.*, 2014; Ridao-López *et al.*, 2012).

Some caution on the interpretation of these differences shall be taken, especially when considering that clinical practice and physicians' decision depend solely on their beliefs (Chandra *et al.*, 2011). Green & Becker (1994), for example, found that variations between geographic areas were not supply-related but instead due to patient's choice on the use of hospital emergency room (Green e Becker, 1994). This finding encloses an important limitation of these studies, SAV may not necessarily mean variation in supply-related factors but instead that other differences in population-related characteristics are not being controlled by standardization and effectively excluded (Folland *et al.*, 2013; Green e Becker, 1994; Ridao-López *et al.*, 2012).

2.1.1. Methods to analyze SAV

According to Volinn, Diehr, Ciol, & Loeser (1994), seven questions should be addresses when assessing variations in medical practice in order to provide useful and relevant information to physicians and policy makers working in this area: 1) What events are to be analysed? 2) What geographic units are to be analysed? 3) How good is the data? 4) Are differences in rates due to chance alone? 5) Are high rates too high? 6) How is a geographic variation to be explained? 7) What is the role of presentation style in explaining geographic variation?

As has been previously mentioned, variations can be addressed taking a geographical perspective or a hospital/provider perspective. When studying the problem on a geographical point of view, i.e., the geographical area where the person lives, it is assumed that the place of residence may influence the access to healthcare. This perspective addresses questions on equity in access or overtreatment driven by excess of supply when the latter depends on the population. When the problem is addressed on a hospital/provider perspective differences may arise from specific characteristics of the providers and organizations. This perspective answers the question whether the geography or the hospital/provider where a person is treated in may influence his/her health outcomes (Bernal-Delgado *et al.*, 2015).

In the case of geographical perspective, the choice of geographic units is not indifferent in the assessment of variations (Srivastava *et al.*, 2014). Analyzing variations in larger areas will result in less variation than when analyzing smaller areas. This happens because, statistically, larger areas utilization rates will be closer to national average than utilization rates of smaller geographical areas (Baicker *et al.*, 2004; Chandra *et al.*, 2011; Fisher *et al.*, 2003; Srivastava *et al.*, 2014) resulting in the smoothing of variations. The choice of the geographical unit shall take this statistical issue in consideration as well as the meaningful geographical unit that is intended to study (Ridao-López *et al.*, 2012).

A limitation from geographic perspective is the possibility that patients receive healthcare in a different geographical area other than that of residence. In these cases, healthcare utilization rate is not truly linked to the place of residence and in many cases it cannot be disentangled in the data and information is lost (Srivastava *et al.*, 2014).

The second step in the analysis is to define which healthcare provision is going to be studied.

Variations of medical practice are less noticeable in interventions with clear reasons to be performed. Some interventions are prone to subjective judgment and thus derive into variations between regions, hospitals, schools of thought, etc. This is called the gray area of medicine (Chandra *et al.*, 2011; Phelps e Mooney, 1993). According to Chandra et al., 2011 there are three attributes for this gray area: scarcity of clinical guidelines, the small probability of harm to patients and idiosyncrasy of clinical benefit.

Researchers on SAV have developed specific frameworks to distinguish the healthcare provision into categories. Although using different names, both Wennberg, Fisher, & Skinner (2003) and ECHO (2014) have similarly identified three different categories. The first category named *Effective Care* or *Evidence-Based Effective Care* would specify healthcare that has proven to benefit all or almost all patients receiving it. If unwarranted variations are identified in this scope of healthcare, it means that somewhere there is a failure in the delivery of needed care and this healthcare provision is being underused. The second category would identify *Preference-Sensitive Care* or *Unclear Benefit-Harm Balance in Non-Elegible Patients.* This is the type of healthcare where clinical benefit is not certain depending on patient characteristics, i.e., patients are not in the clear-cut categories to benefit from the therapy, but they are not in the category that would not benefit at all. Clinical benefits are uncertain as well as the exposure to certain risks. At last, the third category is reserved to the *Supply-Sensitive Care* or *Lower-Value Care*. It distinguishes healthcare whose effectiveness is not clear and where supply may influence utilization rates. In these cases, variations are larger due to availability of healthcare resources or financial incentives. A reference rate would be settled at the rate which additional care will not result in better clinical results but most of the times clinical evidence is lacking (ECHO, 2014; Srivastava *et al.*, 2014; Wennberg *et al.*, 2003).

From these defined frameworks, lessons on the importance of clinical evidence for the reductions of variations are learned. Procedures and activities with good clinical evidence affect clinical practice by harmonizing them and thus reducing variations (Srivastava *et al.*, 2014). Researchers may choose to select a list of procedures and compare them in terms of variability or focus on a specific category of healthcare.

Having defined the point of view of the analysis (population/geographic vs hospital/provider specific) and which procedures to study, the last step on the analysis of unwarranted variations is to define how to quantify variations.

The first common step is the standardization of utilization rates of different geographical areas to accommodate for the differences in population structures and characteristics that may affect crude utilization rates, i.e., if the scope of analysis is a procedure that is more common in women than in men and there is a population with a bigger proportion of women it is expected that the rates of the procedure in that population are greater than in the population with lower proportion of women. Age and sex are the characteristics more commonly used. In the standardization process, corrections on these differences are performed and rates are standardized to assume a similar population structure between geographic areas.

Standardization can be done by two processes: direct or indirect standardization.

Direct standardization computes the expected rates of geographic areas had this specific area had the structure of the overall population (sum of the population of all geographic areas considered) (Atlas de Variaciones en la Práctica Médica; Curtin e Klein, 1995; Pan-American Health Organization, 2002). This method defines what the rate would be if the distribution of the geographic area was like the national distribution (or the distribution of the sum of all geographic areas).

If standardizing for age group (i) and sex (j), mathematically, direct standardization for each geographic area is given by:

$$\sum_{\substack{i=1 \\ j=1}}^{I,J} crude\ rate_{i,j} \times \frac{Procedures_{i,j}}{Total\ Population}$$

Sources: Curtin and Klein 1995; "Standardization: A Classic Epidemiological Method for the Comparison of Rates." 2002; Atlas de Variaciones en la Práctica Médica, n.d.

Indirect standardization computes the expected value for each geographic area given the national distribution rates for each group class considered. It represents the expected utilization if the population of the geographic unit had utilization patterns similar to the overall population considered (sum of all geographic units) (Atlas de Variaciones en la Práctica Médica, [s.d.]; Curtin e Klein, 1995; Pan-American Health Organization, 2002).

Mathematically, indirect standardization for each geographic unit is given by:

$$\sum_{\substack{i=1\\j=1}}^{I,J} No.procedures\ in\ the\ geographical\ unit_{i,j} \times \frac{Total\ no.procedures_{i,j}}{Total\ no.procedures}$$

Sources: Curtin and Klein 1995; “Standardization: A Classic Epidemiological Method for the Comparison of Rates.” 2002; Atlas de Variaciones en la Práctica Médica, n.d.

Having the utilization rates standardized for the population characteristics, researchers have commonly used the Coefficient of Variation (CV), the Variation Ratio (VR) or more sophisticated techniques such as Systematic Component of Variation (SCV) or the Empirical Bayes (EB) to measure SAV (Folland *et al.*, 2013; Phelps e Mooney, 1993; Ridao-López *et al.*, 2012).

CV, computed as the ratio between standard deviation and the mean, will adjust utilization rates to their relative sizes, i.e., when comparing utilization rates of different procedures, variation in the utilization rates of these procedures can now be compared (Folland *et al.*, 2013; Phelps e Mooney, 1993). When using this method, low variation is set between 0,1 and 0,15 while high variation is defined as a CV around 0,5 or above (Phelps e Mooney, 1993). Excess use of a procedure is estimated as the difference between the utilization rate at a specific geographic unit in study and the average (Phelps e Mooney, 1993).

The VR, used for example in Ridao-López et al. (2012), is computed as the ratio between utilization rates of 95%-5%, 75%-25% percentiles or any other percentiles of interest for the researchers. Excess procedures are then computed as the difference between utilization rates of geographic units and the 5% or 25% percentiles (Ridao-López *et al.*, 2012).

SCV and EB provide much more sophisticated techniques to address SAV, though. In these methods, variations related to patient characteristics or with illness severity are removed and only factors related to clinical practice, supply and demand are left to be analyzed (Folland *et al.*, 2013; McPherson *et al.*, 1982; Ridao-López *et al.*, 2012; Shwartz *et al.*, 1994). Both methods depart from the idea that SAV are the sum of two distinct variations: 1) difference across areas in their “true” rates of utilization; 2) variations within areas that is random and are observed around the true utilization rate of the geographical area (McPherson *et al.*, 1982; Shwartz *et al.*, 1994). Variation not attributable to randomness that is recurrent and systematic is then quantified. The SCV considers not the actual utilization rate but the number of procedures that occur relative to the number that are expected given the population characteristics of that area. EB provides an alternative to the

SCV method by giving some weight to the observed rate of the different areas. It results in a weighted average where areas with more reliable observed rates – areas with more observations – have greater weights (McPherson *et al.*, 1982; Shwartz *et al.*, 1994).

If O_i is the observed rate for geographic unit i, and E_i the expected rate given population characteristics, and k is the number of geographic units in analysis, the SCV is given by:

$$SCV = \sum_{i=1}^{k} \frac{\left(\frac{O_i - E_i}{E_i}\right)^2}{k} - \frac{\sum_{i=1}^{k} \frac{1}{E_i}}{k}$$

Sources: McPherson et al. 1982; Shwartz et al. 1994

The EB estimation is then given by:

$$EB = \frac{\sum_{i=1}^{k}\left[\left[\frac{O_i - E_i}{E_i}\right]^2 - \frac{O_i}{(E_i)^2}\right] w_i}{\sum_{i=1}^{k} w_i} \quad \text{and} \quad w_i = \sum_{i=1}^{k} \frac{1}{2\left[EB - \frac{O_i}{E_i^2}\right]}$$

Sources: McPherson et al. 1982; Shwartz et al. 1994

The rationale of these indicators is that values that do not differ from 0 reveal small systematic variation and thus variations are random and given by chance. However how much variation is too much is still subjective. More important is the measurement of excess practice or poor-quality practice that can be avoided and consequently reduce costs.

Poor quality practice has been measured through the identification of failures of care delivery such as adverse events and worse clinical outcomes. Excess practice or overtreatment can be measured by identifying an optimal level of practice given the populations characteristics and measure the difference between this optimal level and the observed level of healthcare provision as in Ridao-López et al. (2012). This leads to the possibility of estimating how much money can be saved or better allocated.

2.1.2. Main Findings on SAV

Peiró & Maynard (2015) summarize the vast literature on SAV in core findings: 1) there is in fact systematic geographic variations in clinical practice that are in their own words substantial, pervasive and persistent over time; 2) patient’s characteristics standardization – by risk-adjustment and health status – does not eliminate the variations identified; 3) no correlation is found between healthcare utilization or spending and better clinical outcomes. (Peiró e Maynard, 2015).

In 1977, Wennberg & Fowler results showed that variations across geographic units could not be explained by population preferences and behaviors (Wennberg e Fowler, 1977), while in 1984, Wennberg presented a work that pointed out the great disparity in the rates of medical treatments, diagnostic tests and surgical procedures in Hospital markets (Wennberg, 1984).

In 1993, non-surgical hospital admissions were identified by Phelps & Mooney as having higher variability than other healthcare services suggesting that there are healthcare services more prone to variability than others (Phelps e Mooney, 1993). The same work revealed that differences in disease incidence and socioeconomic factors were shown to not explain more than a small fraction of the variability found between regions (Phelps e Mooney, 1993).

In 1998, Skinner & Wennberg give a step further by addressing not procedure rates but healthcare spending and clinical outcomes. In this research, regional survival rates were not correlated with more intensive healthcare spending (Skinner, J e Wennberg, 1998).

This study is complemented by the fact that patients tend to prefer less intensive treatments, i.e., not only are healthcare systems increasing healthcare spending without clinical benefits but also physician's choice of treatment is not matching patients' preferences (Skinner, J e Wennberg, 1998).

Again on consumer behavior, Yasaitis, Bynum, & Skinner (2013) identify some differences driven by health status and race but these factors appeared to explain very little of the differences between frequency of office visits (Yasaitis *et al.*, 2013).

In 2015, a work by Ralston, Harrison, Wasserman, & Goodman showed that even after taking medical complexity in account patterns of utilization in children healthcare presented wide variations suggesting that promotion of guidelines and best practices is a need (Ralston *et al.*, 2015).

Regarding the sources of variations, there is no consensus between authors. Zuckerman, Waidmann, Berenson, & Hadley (2010), for example, argue that variation is a result of population disease burden (Zuckerman *et al.*, 2010). In 2014, the work of Gusmano et al. (2014) identified that part of the variations in revascularization rates in France were in fact driven by population-based differences in incidence and mortality of heart attack which were linked to socioeconomic status and risk factors (Gusmano *et al.*, 2014). This would mean that, in some cases, there are, in fact, differences that can be addressed to populations and explain geographical variations in healthcare services. Other authors suggest that these differences in prevalence that are explaining geographic variations are themselves endogenous across regions (Chandra *et al.*, 2011; Song *et al.*, 2010; Welch *et al.*, 2011).

Nevertheless, the extended work of OECD on 13 countries revealed that geographic variation in healthcare use persists after considering differences in demographic structure of the regions showing that this is a phenomenon common across countries (Srivastava *et al.*, 2014).

Broadly speaking research on this topic has revealed that some variations are too large and cannot be explained alone by illness severity or patient preferences (Anthony *et al.*, 2009; Appleby *et al.*, 2011; Baicker *et al.*, 2004; Chandra *et al.*, 2011; Corallo *et al.*, 2014; Hart e Holmstrom, 2010; IOM, 2013; Peiró e Maynard, 2015; Sundmacher e Busse, 2014; Wennberg, 2002). There is evidence that populations exposed to higher rates of use of services do not have longer life expectancy (Peiró e Maynard, 2015; Sirovich *et al.*, 2006; Wennberg, 2002), on the opposite, it seems that overusing health services is not producing better results but instead wasting resources that could be better allocated (Peiró e Maynard, 2015; Sirovich *et al.*, 2006; Srivastava *et al.*, 2014; Wennberg, 2002).

Several researchers have tried to address the reasoning behind these geographic variations other than population-based, focusing on market and supply factors mainly.

Regarding healthcare market, in the last decades there has been significant changes with the modification of morbidity patterns towards chronic conditions and the rapid adoption of technological innovation (Peiró e Maynard, 2015). But it is the misuse of new technologies that are, probably, promoting the over and underuse of healthcare services increasing geographic variations (Peiró e Maynard, 2015).

This is corroborated by the works of several authors that report that technological diffusion, availability of specialists, local training framework that differ across regions are contribution to variations in clinical practice (Birkmeyer *et al.*, 2013). In fact, this is expected when physicians themselves are not consistent on their opinions on the value or need of a therapy (Wennberg e Fowler, 1977). Wennberg (1984) suggests that variations are related to the uncertainty of clinical benefits of therapies that creates disparities in clinical practices (Folland *et al.*, 2013; Wennberg, Barnes e Zubkoff, 1982; Wennberg, 1984). Another underlying factor is the uncertainty of the diagnosis. While not sure on the diagnosis, physicians' practice may also vary perpetuating different clinical practices along the years (Barros, 2009; Grytten, Monkerud e Sorensen, 2012; Wennberg *et al.*, 1982).

Other underlying factor for geographical variations may be related to financial incentives and regulatory factors differing across regions (Birkmeyer *et al.*, 2013). In any case, whether differences are related to clinical practice and beliefs on the benefits of treatments or to economic incentives, these findings reveal that physicians are able to influence healthcare patterns and promote utilization above the clinical need (Birkmeyer *et al.*, 2013; Fisher *et al.*, 2003; Srivastava *et al.*, 2014).

Healthcare services' use constraints are usually applicable to consumers though co-payments or other usage limitation. However, according to the evidence found, these policies may not be the most effective, since a great deal of variability and overuse lies in the supply side (Wennberg e Fowler, 1977).

2.1.3. Supply-induced demand

If physicians are, in fact, able to promote healthcare use beyond clinical needs then SAV introduces the notion that there are geographic variations that persist even after standardization of physicians training and that physicians can have different clinical practices whether they are justified for different beliefs or lack of clinical evidence that generate different treatment patterns. What if those differences are generated not on the best interest of patients? i.e. What if physicians could choose not optimally clinical treatment in their own favor? (Chandra *et al.*, 2011; Fisher *et al.*, 2003; Johnson, 2014).

Physicians' influence on demand arises from asymmetric information, a well-known market failure. This generates the concept of *agency* i.e., having more information than patients on the treatments and quantities of those treatments, physicians act in the best interest of patients as patients would act if they have themselves the complete information (Chandra *et al.*, 2011; Folland *et al.*, 2013; Johnson, 2014). But as physicians work as agents for patients this means that while they are expected to act in the best interest of patients, physicians could also neglect patient's best interest for their own profit or not (Folland *et al.*, 2013).

These hypothesis have originated the theories of Physician-Induced Demand (PID) or Supply-Induced Demand (SID). Evans (1974) hypothesized physicians, having more information than patients, could influence the demand for their own services (Evans, 1974). The definition of this concept is provided by T. McGuire (2000): *Physician-induced demand exists when the physician influences a patient's demand for care against the physician's interpretation of the best interests of the patient* (McGuire, T., 2000). T. McGuire (2000) goes further by explaining that economic incentives, such as quantity rewarding payment schemes, reduction of fees or increased supply may lead physicians to influence patient's demand curve accommodating physicians own interests (McGuire, T., 2000).

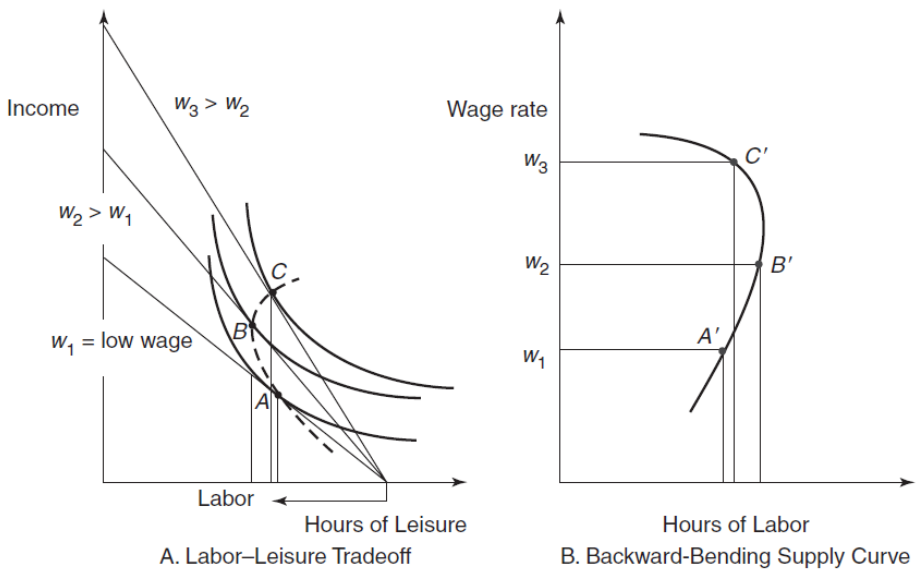
It is worth mention that the definition of SID clearly excludes physicians' choices that are taken in the patients' best interest, i.e., it is expected that physicians choose treatments that increments the utility of patients. These treatments are not the scope of this definition. Instead, SID theory studies the extra healthcare services which will not increment patients' utility but instead are taken in the physician own interest (Chandra *et al.*, 2011).

SID theory has been used as one of the justifications for the increase in health expenditure which explains the importance given to this topic (Cromwell e Mitchell, 1986; Evans, 1974; McGuire, A., Henderson e Mooney, 1988; Newhouse, 1992; Rice, 1983).

According to SID theory, individuals respond to incentives and physicians are no exception (Folland *et al.*, 2013). Physicians will thus behave in order to maximize their utilities: physicians value net income and leisure and dislike inducing patient demand (Folland *et al.*, 2013; Johnson, 2014; McGuire, T., 2000).

The "solution" of the "problem" lies in the trade-off between income and leisure and in the trade-off between income and inducement (Folland *et al.*, 2013). Considering income-leisure trade-off (Figure 2), for different wage rates there are different budget constraints. Physicians will choose their optimal point in each income line provided the additional income they get for each hour of leisure they give up (Folland *et al.*, 2013; Johnson, 2014; McGuire, T., 2000).

Figure 2: Graphic representation of Income-Leisure trade-off



Source: Folland *et al.*, 2013. p. 303

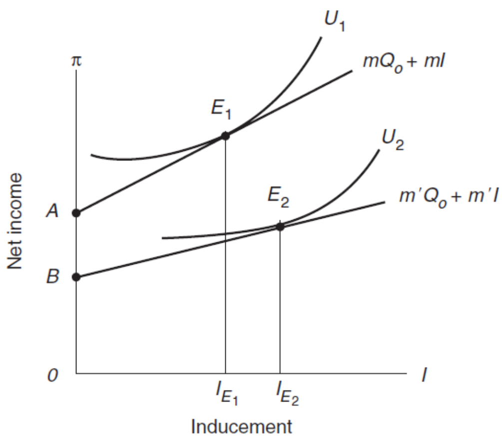
Panel B on the same Figure, represents the supply of labor curve showing how much work physicians are willing to take for each wage rate (Folland *et al.*, 2013). From point A' to B', the increase in wage rate motivates the physician to work more and substitute leisure for work and get higher income levels. But from point B' to C', physician's income level is high enough to them to will to take some more leisure time instead of labor (Folland *et al.*, 2013). Understanding this trade-off theory is crucial to figure how physicians are willing to work (Folland *et al.*, 2013). If physicians wage goes down for whatever reason from C' to

B’, then, according to this model, physician will compensate the lost income by working more (Folland *et al.*, 2013; Johnson, 2014; McGuire, T., 2000).

The next trade-off (Figure 3) considers net income and inducement. It is assumed that physicians dislike inducing demand on patients and each time they induce demand there is a disutility in this action (Folland *et al.*, 2013). Lost utility by inducing demand will need to be compensated by the extra utility net income provides (Folland *et al.*, 2013; Johnson, 2014; McGuire, T., 2000).

In A, physician is not inducing demand, but his utility optimizing behavior will shift his inducement practice to point E_1 as this behavior will increment his utility (U_1). Supposing his wage rate declines, his net income-inducement curve will get flatter and shift downwards. In this case, his no-inducement point would be B and his optimizing behavior will shift his inducement practice to E_2 which means a wage rate decrease will increase inducement practice: $E_2 > E_1$ or by other words, a change in physician fee will affect the quantities physician supply (Folland *et al.*, 2013; Johnson, 2014; McGuire, T., 2000).

Figure 3: Graphic representation of Income-Inducement trade-off



Source: Folland *et al.*, 2013. p. 304

In summary, the model considers two effects: the substitution effect that addresses whether it is worth to move from leisure to more labor (by induced-demand) and an income effect that addresses whether changes in income would change inducement patterns (Folland *et al.*, 2013). When income effect is greater than the negative substitution effect caused by inducement, then physicians will have an incentive to induce demand (Folland *et al.*, 2013; McGuire, T. e Pauly, 1991).

Criticism to this theory justifies that theory models are not correct because they do not address patient's differences nor do they address differences in practice styles, practice environment among other factors (Johnson, 2014). Nevertheless, SAV evidence has demonstrated that demand characteristics and differences in populations are not explaining much of variations in clinical practice across regions and potentially across physicians (Folland *et al.*, 2013; Folland e Stano, 1990).

As for the practice style mentioned, it may be viewed as an incomplete knowledge of physicians on the true value of the treatments and alternatives that is usually corrected with education and monitoring (Folland *et al.*, 2013). While promoting differences in supply, physicians that are not well-informed on the treatments they are providing, may not be intentionally inducing demand (Folland *et al.*, 2013).

But controversies to SID theory are more pronounced when perfect agency or target income hypothesis are considered.

In the first, researchers state that one does not need to assume SID to have an increment in quantity of care following and increment in supply. The traditional supply-demand theory would predict that in a simpler manner (Carlsen e Grytten, 1998; Folland *et al.*, 2013; McGuire, T., 2000). To address this question Reinhardt (1985) has proposed the "fee test" where inducement can only be invoked if physicians' fees increased to levels above the initial accommodating not only supply increase but also increased demand from SID (Folland *et al.*, 2013; Reinhardt, 1985). Including quality in the model as proposed by Feldman & Sloan (1988) can have the same effect, though. In this case, physicians would react to increased supply with an increase in quality which in turn could mean higher prices excluding once more the SID from potential explanation to increased quantity and prices (Feldman e Sloan, 1988; Folland *et al.*, 2013).

As well, responses to fee reductions may not be attributable to SID as the effects after an imposed price reduction on the traditional supply-demand model are similar (McGuire, T., 2000). In alternative, analyses of income-shocks will differentiate maximizing suppliers from inducers (Johnson, 2014). Studies using informed and non-informed patients can also be used to address differences in physician behavior (Johnson, 2014).

The point is that agency relationship and asymmetry of information may violate the hypothesis assumed by neo-classical theory that supply and demand are independent thus supply-demand theory may not be correctly addressed in these cases (McGuire, T., 2000).

The target income hypothesis considers that physicians set a target income and thus their behavior is modeled by this target (Folland *et al.*, 2013). This would exclude any SID theory but instead that physicians adjust prices and quantities to achieve their target not even considering extra income. Defenders of this theory include Rizzo & Blumenthal

(1996) and Rizzo & Zeckhauser (2003) (Rizzo e Blumenthal, 1996; Rizzo e Zeckhauser, 2003).

The literature on inducement is as conflicting as the theory. Counter arguments are especially proliferous because of the difficulty to test SID empirically (Carlsen e Grytten, 1998).

According to Carlsen & Grytten (1998) initial studies on the topic support the SID theory while later studies reject it (Carlsen e Grytten, 1998; Dranove e Wehner, 1994; Escarce, 1992; Evans, 1974; Fuchs, 1978; Redisch, Gabel e Blaxall, 1981; Stano, 1985). The reason behind this may be technical. Initial studies were performed on aggregate data and looked for an availability effect, i.e., that an increase in supply would lead to an increase in the services provided by physicians (Chandra *et al.*, 2011; Cromwell e Mitchell, 1986; Feldman e Sloan, 1988; Fuchs, 1978). Later studies, have addressed the question differently by examining exogenous demand shocks and analyzing changes in utilization after reductions in physicians' payment (Chandra *et al.*, 2011; Dranove e Wehner, 1994; Gruber e Owings, 1996; Nguyen e Derrick, 1997; Rice, 1983; Rossiter e Wilensky, 1984; Yip, 1998).

There are several studies that show that physicians do respond to economic incentives such as reimbursement schemes especially if they promote quantities (Folland *et al.*, 2013). Initial work of Evans (1974) addressed exactly this question through the analysis of increased demand in Canada and United States. Rice (1983) found as well evidence of SID when finding greater quantities of surgeries and laboratory tests (Rice, 1983).

Physicians paid on a fee-for-service have more incentive to supply more healthcare services than physicians on capitation schemes (Nassiri e Rochaix, 2006; Quast, Sappington e Shenkman, 2008). The study of Iversen (2004) has demonstrated that physicians with shorter lists of patients increase their lists to compensate the reduced income and study by Rizzo & Zeckhauser (2003) suggest that physicians pursue income targets (Iversen, 2004; Rizzo e Zeckhauser, 2003). The work of T. McGuire & Pauly (1991) also concludes that when income-effects are in place, physicians will seek a target income (McGuire, T. e Pauly, 1991). Another study by Ho & Pakes (2011) refers that physicians choose treatment options they have incentives to (Ho e Pakes, 2011).

Regarding changing in fees paid to physicians, Nguyen & Derrick (1997) have shown that when facing reduction in fees, physicians increase volume to generate additional income that was lost (Nguyen e Derrick, 1997). The same evidence is found by Yip (1998) in whose study volumes of Coronary Artery Bypass Grafting (CABG) increased more for physicians whose income reduced more (Yip, 1998).

Regarding increase in competition and the reaction of physicians, Gruber & Owings (1996) have found evidence that increase in competition has lead physicians to promote higher paid services as C-sections (Gruber e Owings, 1996). Fuchs (1978) estimates that a 10 percent increase in the ratio between surgeons and patients will lead to an increase of 3% in per capita utilization showing that increased competition, increase healthcare services volume (Fuchs, 1978). Cromwell & Mitchell (1986) identified areas with more surgeons as areas with higher utilization rates and higher fees (Cromwell e Mitchell, 1986). Carlsen & Grytten (1998) on other side do not find clear evidence than competition reflects SID or if other factors are the justification (Carlsen e Grytten, 1998).

As well, services that are more profitable than others, such as specific surgery settings have evidence to be more used than others (Fuchs, 1978; Plotzke e Courtemanche, 2011). Grytten, Carlsen, & Sørensen (1995) find evidence on inducement for laboratory tests but not for physician visits which may be explained by these differences in profitability (Grytten *et al.*, 1995).

Although Rossiter & Wilensky (1984) find evidence of SID, the magnitude of the effect is considered small and is only significant for optional treatments (Rossiter e Wilensky, 1984).

Several critiques to the studies presented above have proliferated across literature especially on methodological questions. These studies are thought to favor inducement evidence by not excluding accurately several factors such as supply shocks or by using variables that are thought to be endogenous as average coinsurance rates (Gruber e Owings, 1996; Johnson, 2014). Dranove & Wehner (1994) method, for example, has suggested that physicians were influential on the number of births which is something that is certainly not expected and was thought to be caused by a misspecification in the model that had been used before in the work of Cromwell & Mitchell (1986) (Dranove e Wehner, 1994).

But apart from methodological critiques, other authors did not find evidence of SID. This is the case for early study from Feldman & Sloan (1988). While they agree SID may be present for some profitable healthcare services, they state that this behavior is much lower than what was suggested by other authors (Feldman e Sloan, 1988). Carlsen & Grytten (1998) did not find evidence of SID in visits or in laboratory tests either (Carlsen e Grytten, 1998).

Apart from understanding whether there is or there is not SID, Labelle, Stoddart, & Rice (1994) work on a different question: whether the inducement of demand affects the health of patients. In their work, they included clinical effectiveness and effectiveness of agency relationship. They conclude that conceptual frameworks on inducement are too limited to develop policies on health service utilization (Labelle *et al.*, 1994). Another study developed in Belgium, a country where healthcare financing – based on fee per procedure – promotes SID, was developed to understand how SID could affect patients' behavior, namely willingness-to-pay. Results suggested that SID did not affect significantly patients' behavior (Voorde, Van, Doorslaer, Van e Schokkaert, 2001).

2.1.4. Addressing geographical variations

An initial thought on how to solve geographical variations is clearly the need for guidelines. As Folland et al. (2013) suggests, the lack of consensus on the clinical benefits of several medical procedures generates disparities on how patients are treated (Folland *et al.*, 2013). The different beliefs of physicians, patterns of practice, different treatment choices, etc. are justifications for variations in clinical practice especially when there are incentives for these variations. Those variations could be reduced if clinical practice was homogenized with better scientific evidence and clinical guidelines (Birkmeyer *et al.*, 2013; Folland *et al.*, 2013).

Physicians' training and information programs may also promote better behaviors and influence practice styles (Folland *et al.*, 2013; Wennberg e Fowler, 1977). The work of Wennberg & Fowler (1977) showed that tonsillectomy's rates have been significantly affected by an informational program that was established in England (Folland *et al.*, 2013; Wennberg e Fowler, 1977). Another information program directed to physicians reduced pelvimetries performed without need (Chassin e McCue, 1986; Folland *et al.*, 2013). In Canada, a review program reduced unjustified hysterectomies (Dyck *et al.*, 1977; Folland *et al.*, 2013).

Wennberg (1984) has suggested that the first step towards correct decisions on resource allocation is the monitorization and information on regional performances (Wennberg, 1984).

Many similar and other policies are identified in the OECD report and can be used to incentivize the correct utilization of healthcare services more than to reduce variations. These were described as soft-touch policies and include: the public reporting of geographical variations as to create awareness of these variations on stakeholders and promote actions; setting targets at the regional levels to promote appropriate use of healthcare services; re-allocation of resources to correct for over or under treatment that is thought to be influenced by supply; implementation of clinical guidelines that homogenizes clinical practice; provider-level reporting and feedback; changes in payment systems to correct of over or under treatment; measurement of health outcomes; utilization of decision aids for patients promoting informed decisions and responsiveness to patients' preferences (Srivastava *et al.*, 2014).

In Spain, after the development of monitorization of indicators for misuse of C-sections the rates have reduced or stabilized across different regions and the establishment of clinical guidelines in the United Kingdom have kept C-section rates at an average level below OECD level with low geographical variability (Alvarez-Bartolomé e Gogorcena-Aoiz, 2014; Farebrother, 2014).

In knee arthroscopy, Patient Reported Outcomes Measures (PROMs) are being developed to determine the benefits of these interventions to patients (Farebrother, 2014).

Although a lot of research work has been done regarding the identification of medical practice variation, there is not much work on the evolution of this phenomenon. The literature review of Groenewegen and Westert (2004) on works analyzing the evolution of medical practice variation concludes that in general, variations in medical practice have been reducing over the years after implementation of guidelines and utilization of data for management of variations and regulation of physician behavior.

2.2. Efficiency

Efficiency and waste have been major topics of concern along the times but have become hot topics in the last two decades. Scarcity of resources, financial crisis and the necessity to become more and more efficient has increased the focus on these topics on the whole society and on healthcare system as well.

Efficiency has been described in the literature as the best use of resources in production (Hollingsworth, Dawson e Maniadakis, 1999) or the extent to which objectives are achieved in relation to the resources consumed (Jacobs, Smith e Street, 2006). Farrell (1957), one of the reference authors on the topic has defined efficiency as the production of as many possible outputs given an available set of inputs (Farrell, 1957). All these definitions, however, imply the use of available resources in the best possible manner, whether it is more output, better quality output or objectives achievement. Additionally, efficiency definitions can be divided in three types of efficiency: 1) technical, allocative and cost efficiency. While the definition of technical efficiency is virtually consensual – maximization of outputs for a given level of inputs or minimization of inputs for a given level of outputs – the definitions of allocative and cost efficiency differ in some points (Cylus *et al.*, 2016; Farrell, 1957; Street *et al.*, 2011). Street et al. (2011) define allocative efficiency as the choice of the appropriate mix of inputs and outputs to maximize utility and cost efficiency as the minimization of costs for any given output level, Farrell (1957) defines allocative efficiency as the correct choice of inputs, given their prices, to maximize output (Cylus *et al.*, 2016; Farrell, 1957; Street *et al.*, 2011).

As Cylus, Papanicolas, & Smith (2016) have identified there are several issues regarding inefficient health systems: i) patients not receiving best possible care; ii) inefficient treatment potentiate the denial of treatment to other patients that could have benefited; iii) inefficiency in health system mean opportunity costs in other economic sectors such as education; iv) inefficient health system may harm social solidarity and social welfare (Cylus *et al.*, 2016).

Failures of care delivery, failures of care coordination, overtreatment, administrative complexity, pricing failures, fraud and abuse are identified by Berwick & Hackbarth (2012) as sources of waste in healthcare services delivery.

Failures of care delivery are related to poorly executed services or performance of non-best practices, for example, which is intimately related to medical practice variations. This kind of failures leads to patient injuries, adverse events and worse clinical outcomes that incur in not only additional costs for hospitals but also represent patient discomfort, inconvenience and potential loss of income (Bernal-Delgado, Martínez-Lizaga e Ridao-López, 2011; Berwick e Hackbarth, 2012; Jackson *et al.*, 2011).

Methods for identifying these failures often try to assess the occurrence of adverse events and to quantify the excess costs they produce. Ideally, adverse events could be directly identified in the comorbidities of an inpatient stay and thus easily quantified, however, this information is not always available and this lack of information may exclude about 41% of the cases considered failures of care delivery (Jackson *et al.*, 2006).

Bernal-Delgado et al. (2011) measure the excess length of stay driven by the occurrence of thromboembolism after surgery, assuming the increase in the length of stay is a proxy of the excess costs incurred. Different methods such as linear regression, matched controls, cluster effects controls, etc. were used to understand how thromboembolism affects the length of stay. Age, gender, comorbidities, severity and hospital characteristics were also considered.

Limitations on these methods are related to the quality of the data on secondary diagnoses and coding heterogeneity across institutions (Bernal-Delgado *et al.*, 2011). Further difficulties arise on identifying if specific diagnoses codes were already present at hospital admission or if they were hospital-acquired (Jackson *et al.*, 2006, 2011).

Ridao-López et al., (2012) assessed the economic losses incurred by variability in medical practice. Excess number of cases was multiplied by a unitary cost computed per hospital (the current spending of the hospital divided by the total sum of All-Patient Refined Diagnostic Related Groups – APR-DRG – cost weights) and multiplied by the cost-weight associated with each patient APR-DRG.

When variability of medical practice exists, it can be the case that overtreatment is in place. Overtreatment refers to subject patients to care that, according to sound science and the patients' own preferences cannot possibly help them (Berwick e Hackbarth, 2012).

Measurements of unnecessary spending depend, largely, in which interventions are considered in the analyses and the definition of the optimum rate of those interventions.

Common to both overtreatment and failures of care delivery are several limitations: i) coding heterogeneity; ii) definition of best practice; iii) definition of geographic areas; iv) and cost measurement, that need to be addressed when assessing variations in medical practice.

Given the need to reduce costs and expenditures, analysis of efficiency may be a powerful tool for helping to create cost-containing policies and to understand where resources are being wasted.

Measuring efficiency is an essential activity when information on how good a specific service is, is required. Measurement isolates the good from the bad according to a specific consensus and should aim to provide relevant information to those being measured (Papanicolas e Smith, 2013; Smith, Mossialos e Papanicolas, 2008).

Considering the definition of technical efficiency – maximization of output for a given level of inputs – and that inefficient treatment potentiate the denial of treatment to other patients that could have benefited, variations in medical practice generate inefficiencies through the delivery of care that: 1) is non-optimal; 2) uses resources that could have been better allocated.

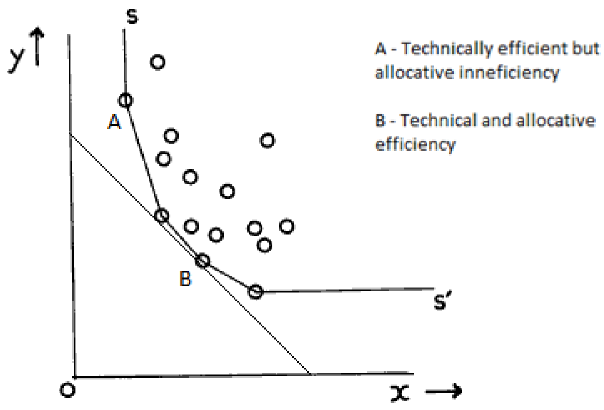
Comparisons between practitioners, institutions and, ultimately, health systems provide information to physicians, healthcare institutions, governments and populations on the state of their health system, how good the medical practice is and how it compares to their direct competitors or peers (García-Altés *et al.*, 2006; Smith *et al.*, 2008). The information provided by efficiency measurement creates the opportunity for improvement and accountability of practice being also a measure to justify decision making (Smith *et al.*, 2008).

2.2.1. Methods to estimate technical efficiency

Graphically, technical efficiency means that there is an isoquant in which different combinations of inputs produce the same quantity of output (Figure 4). Above that isoquant combinations of inputs producing the same quantity are being inefficient because they are not minimizing inputs (Farrell, 1957; Franco e Fortuna, 2003).

Budget constraint, also represented on Figure 4, allows understanding allocative efficiency by expressing the ratio of prices of inputs. Point A may be technically as efficient as point B but requires a greater budget while point B produces the same output at a lower cost level.

Figure 4: Technical and Allocative Efficiency



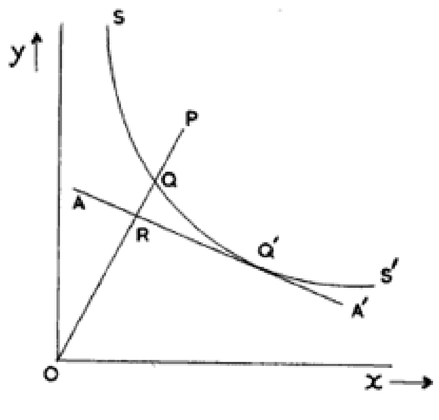
Source: Farrell, 1957. p. 256

To estimate inefficiency, Farrell (1957) proposes the estimation of an isoquant indicating the possible input combinations to produce a certain level of output and the comparison of the observed results with the inputs used and the prices of the inputs.

Looking at Figure 5, suppose isoquant SS' had been estimated. A firm performing at point Q' would be allocative and technically efficient, point Q would be technically efficient but would show allocative inefficiencies and point P would be neither allocative nor technically efficient.

P inefficiency is a sum of technical and allocative inefficiency. P shows technical inefficiency because it requires a combination of more inputs to produce the same as Q and Q'. The distance from the isoquant represents the technical inefficiency of P, and it is the ratio 0Q/0P. Moreover, P shows allocative inefficiency by requiring a greater budget constraint – P does not lay on the AA' line. Allocative inefficiency from P is measured as the distance from the point where the line 0P crosses the isoquant SS' to point R, i.e., the ratio 0R/0Q. Overall the total inefficiency is measured by 0R/0P.

Figure 6: Estimation of inefficiency



Source: Farrell, 1957. p. 254

Although theoretically it seems simple, estimating the isoquant might be complex. Methodologies to assess efficiency have been developed since the work of Farrell but it has been specially challenging to use them given the lack of a common basis for comparison (Busse e Quentin, 2011; Cylus e Pearson, 2016; Jacobs *et al.*, 2006; Sund e Häkkinen, 2016).

The most commonly used frontier methodologies developed following Farrell's work are Data Envelopment Analysis (DEA) and the Stochastic Frontier Analysis (SFA) (Charnes, Cooper e Rhodes, 1979; Hollingsworth *et al.*, 1999; Hollingsworth, 2016; Jacobs *et al.*, 2006; Retzlaff-Roberts, Chang e Rubin, 2004).

DEA is a non-parametric methodology (Franco e Fortuna, 2003; Retzlaff-Roberts *et al.*, 2004) where the production frontier is estimated taking into account all hospital performances included in the analysis and measuring the relative efficiency of each unit.

Developed by Charnes, Cooper, & Rhodes (1979) it assumes constant returns to scale and measures global efficiency. Each unit (hospital for this matter) is compared to the others allowing the identification of the efficient and inefficient hospitals. The production frontier is set around the production frontier of the units considered the most efficient; units lying above that production frontier are considered inefficient. The distance to the production frontier is the inefficiency level (Charnes *et al.*, 1979). This methodology was later developed by Banker, Charnes, & Cooper (1984) to include variable returns to scale hypothesis (Banker *et al.*, 1984). The estimated production frontier is given by:

$$max h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \quad \text{subject to} \quad 1 \geq \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}}, \quad j = 1, \dots, n,$$

$$\text{with} \quad u_r, v_i > 0, \quad i = 1, \dots, m; \quad r = 1, \dots, s$$

Source: Banker, Charnes and Cooper (1984)

Where $y_{rj}, x_{ij} > 0$ are respectively the output and input for decision making unit j.

In summary, DEA will identify the maximum output produced for each set of inputs defining the optimal production function in this way.

SFA is other frontier method commonly used to estimate efficiency. SFA was developed simultaneously by Aigner, Lovell, and Schmidt (1977) and Meeusen and van Den Broeck (1977) and is a parametric methodology where the production function is estimated through a parametric function such as Cobb-Douglas (Coelli, Prasada Rao e Battese, 1998). SFA assumes that the distance between the isoquant and the point where the firm is positioned is a result not only from inefficiencies but also from random factors out of the firm control. This means that the best performance of a firm may suffer random shocks and each firm needs an individual expected frontier. The error term is then defined as a sum of two components, a one-side distribution term that represents the specific inefficiency term and a normal distributed term that represents random fluctuations out of hospital control (Aigner *et al.*, 1977; Coelli *et al.*, 1998; Li e Rosenman, 2001). This stochastic process gives name to the method in study (Farsi e Filippini, 2005; Franco e Fortuna, 2003; Rosko, 2001). The production function estimated by SFA is given by:

$$y_i = f(x_i; \beta) + \varepsilon_i \quad \text{and} \quad \varepsilon_i = u_i + v_i$$

Source: Aigner et al. 1977

where y_i is the output, f is the functional form of the technology, x_i is the vector of inputs and β a vector of parameters to be estimated.

The structure of the error term is the key point of the SFA by being composed by a symmetric disturbance v_i that is independently and identically distributed as $N(0, \sigma_v^2)$ and represents the random fluctuations out of hospital control. u_i is independent of v_i and represents the inefficiencies of the hospital (Aigner *et al.*, 1977; Coelli *et al.*, 1998).

Technical efficiency is defined as the percentage of the production function attained by the hospital and is the ratio of observed output relative to the potential output of the hospital given its input level (Coelli *et al.*, 1998):

$$TE_i = \frac{y_i}{exp(x_i\beta)} = \frac{exp(x_i\beta - u_i)}{exp(x_i\beta)} = exp(-u_i)$$

Source: Coelli et al., 1998

SFA provides an approach that is coherent with the economic theory; it is necessary to define a priori the structure of the technology (Coelli *et al.*, 1998; Jacobs *et al.*, 2006). The functional forms more commonly used are the Cobb-Douglas production function or the Translog model which is a generalization of the previous (Coelli *et al.*, 1998). Although the Cobb-Douglas functional form is simpler than the translog model, it has some restrictive properties such as constant input elasticities, constant returns to scale and the assumption that elasticity of substitution equals one. On the other hand, the translog form does not impose restrictions on the returns to scale nor on substitution possibilities. However, the translog form is highly sensitive to multicollinearity and has problems with degrees of freedom (Coelli *et al.*, 1998). Tests on the results obtained through different functional forms state that inefficiency estimates are not very sensitive to the choice of functional form (Rosko e Mutter, 2008, 2011; Zuckerman, Hadley e Iezzoni, 1994).

SFA allows for modelling and measurement of the error by adding an error component that reflects random shocks out of hospitals control (Aigner *et al.*, 1977; Coelli *et al.*, 1998; Jacobs *et al.*, 2006; Li e Rosenman, 2001). As for the functional form, SFA requires defining a priori the distribution of the one-sided non-random part of the error term which is sometimes arbitrary (Coelli *et al.*, 1998). While there are four available distributions (half normal, truncated-normal, exponential and gamma distribution), the literature, both generic and in healthcare, states that SFA results are robust across different definitions of error distribution (Rosko e Mutter, 2008, 2011; Zuckerman *et al.*, 1994).

Contrarily to DEA, SFA has the limitation of allowing the inclusion of one output only. To overcome this limitation, when applied to multi-product industries, SFA analysis can either introduce a summary measure of output – like an index – or to estimate a cost-function. In this case, the cost-function would be minimized (Jacobs *et al.*, 2006; Rosko e Mutter, 2011). A cost-function allows that multiple outputs can enter the cost-function as separate independent variables (Rosko e Mutter, 2011). Cost-functions may be preferable to production functions since it can be assumed the exogeneity of input prices while this is not always true when using input quantities in a production function (Vitaliano e Toren, 1994). Given the advantages presented, the majority of studies that apply SFA focus on cost-functions (Rosko e Mutter, 2011).

Estimating and comparing efficiency measures is more robust when organizational structures are similar, when quality of the data is assured and when data is collected using the same criteria (Cylus e Pearson, 2016; Medin *et al.*, 2013; Smith *et al.*, 2008; Sund e

Häkkinen, 2016). The complexity of the production process, the diversity of the outputs produced in the different organizational environments and the lack of relevant, reliable and comparable data have made it difficult to measure or compare institutions and healthcare systems (Bojke *et al.*, 2013; Jacobs *et al.*, 2006; Sund e Häkkinen, 2016).

2.2.2. Measuring outputs and inputs

Measuring output is not straightforward. This happens because the true final output – improvement in the health of patients – is difficult to assess in an objective manner in all the patients discharged (Sund e Häkkinen, 2016). The number of patients treated (discharges) or the number of inpatient days are easily measured and quantifiable and represent the production of the hospitals or healthcare systems. Some hospitals can produce more discharges than others, as if they were different factories. Discharges can therefore be thought as a measure of hospital output.

These measures, however, are somehow limited because they do not truly reflect the different combination of resources used to treat the patient, neither the severity nor the complexity, assuming all discharges and all inpatient days are similar. To overcome these limitations some authors have weighted both discharges and inpatient days with the associated case-mix or reference cost (Bojke *et al.*, 2013). Several other studies used Diagnostic Related Groups (DRG) cases to measure outputs (Kittelsen *et al.*, 2009, 2015; Linna *et al.*, 2010; Linna, Hakkinen e Magnussen, 2006; Medin *et al.*, 2011).

Other limitation to be considered when talking about outcome is that outcome may not reflect all production from a hospital or healthcare system. Bojke et al. (2013) included all healthcare services provided to National Health Service (NHS) residents except primary care. If only a part of the activity is considered and all the remaining processes from a hospital are not considered the production function obtained is only partial.

Translating outputs and inputs into monetary units is a way to simplify the estimations and aggregate different outputs and different inputs overcoming the existing diversity (Bojke *et al.*, 2013). These calculations are not always possible however. Several authors combine different output units. Others estimate efficiency with a mix of output units and input prices.

On the inputs side, works on this topic have been using information that includes medical, nursing or total staff, with wage differences taken into account, salaries paid by hospitals, pharmaceutical and other intermediate inputs, operational costs, capital costs, available beds, etc. (Bojke *et al.*, 2013; Coelli *et al.*, 1998; Farsi e Filippini, 2005; Gonçalves e Barros, 2013; Kittelsen *et al.*, 2009, 2015; Kristensen *et al.*, 2012; Linna *et al.*, 2010, 2006; Medin *et al.*, 2011; Schleiniger, 2008; Wang, Zhao e Mahmood, 2006).

Hospitals’ characteristics, such as size, case-mix index, type of hospital, etc., have also been introduced in the analysis of efficiency being thought to possible influence the efficiency of the hospital (Kang, Hong e Park, 2012; Oliveira e Bevan, 2008).

3. Objectives

Identifying, understanding and estimating the consequences of geographical variations in medical practice are the keys to reduce this phenomenon and its associated inefficiencies. The comparison of regions, institutions and healthcare systems provide insights on optimal and non-optimal performances. However, the lack of comparable data has limited these analyses. With the development of international databases new opportunities are created for the comparison and analysis of health systems.

The main objective of this thesis is to contribute to the understanding of variations in medical practice and their implications for the health system. Specific objectives have been organized in three categories: Identification, Causes and Consequences of Variations in Medical Practice. Five concrete objectives have been envisaged due to their contribution to the main objective.

<u>Objective 1</u>: To identify geographic variations in health care utilization in Portugal based on patients' place of residence. A specific set of procedures and health activities are selected based on high-cost, high-volume, policy relevance and data availability. Those include: medical admissions, coronary artery bypass grafting, percutaneous transluminal coronary angioplasty, cardiac catheterization, surgery after hip fracture, knee replacement, knee arthroscopy, caesarean section and hysterectomy.

<u>Objective 2</u>: To quantify the geographical variability and excess use of C-section, an effective procedure but which rates are considered above what is clinically recommended. Additionally, mapping identifies geographical areas which could reduce rates and economic burden of this excess use is also calculated.

<u>Objective 3</u>: To study whether resource availability can explain the excess use of C-sections in Portuguese NHS hospitals and the costs of excess use.

<u>Objective 4</u>: To study Stochastic Frontier Analysis as a methodology for efficiency measurement and to identify previous studies using this methodology in health economics to assess variables used, main results and limitations.

<u>Objective 5</u>: To assess and compare hospital efficiency levels within and between countries using Stochastic Frontier Analysis with both cross-sectional and panel data collected from an international database (ECHO).

Objectives are achieved in the form of research works presented in the Results.

Table **1** identifies the works corresponding to each objective.

Table 1: Research work corresponding to each objective

Objective	Research Work Produced
Identifying the issue and measuring it. (Identification)	
Objective 1 To identify geographic variations in health care utilizations in Portugal	Portugal: Geographic variations in health care *Mateus, C., I. Joaquim, C. Nunes, P. Boto, and L. Campos. 2014. "Portugal: Geographic Variations in Health Care." In Geographic Variations in Health Care: What Do We Know and What Can Be Done To Improve Health System Performance?, 317–42. OECD Health Policy Studies.*
Objective 2 To quantify the geographical variability and excess use of C-section	Potential of geographical variation analysis for realigning providers to value-based care. ECHO case study on lower-value indications of C-section in five European countries *García-Armesto, S., Angulo-Pueyo, E., Martínez-Lizaga, N., Mateus, C., Joaquim, I., Bernal-Delgado, E., on behalf of the ECHO Consortium. 2015. "Potential of geographical variation analysis for realigning providers to value-based care. ECHO case study on lower-value indications of C-section in five European countries." The European Journal of Public Health 25 (suppl 1): 44–51. doi: 10.1093/eurpub/cku224.*
Understanding the reasons for variations in medical practice. (Causes)	
Objective 3 To study resource availability as a reason for excess use	Hospital characteristics and avoidable C-sections: a decade analysis *Submitted to an international journal*
Measuring the inefficiencies resulting from variations in medical practice. (Consequences)	
Objective 4 To study Stochastic Frontier Analysis as a methodology	Estimating hospital efficiency using SFA – Lessons from previous studies *Submitted to na international journal*
Objective 5 To compare hospital efficiency levels within and between countries	Measuring hospital efficiency — comparing four European countries *Mateus, C., I. Joaquim, and C. Nunes. 2015. "Measuring Hospital Efficiency--Comparing Four European Countries." The European Journal of Public Health 25 (suppl 1): 52–58. doi:10.1093/eurpub/cku222.*

4. Data Sources

The European Collaboration for Health Optimization (ECHO) (www.echo-health.eu) is a common effort of five European countries that enabled the construction of an international database combining not only clinical data but also socioeconomic data for research purposes. This thesis and its results were dependent on this database.

The data included in this database can be divided in three categories:

- Inpatient hospital episodes
- Hospital resources
- Socioeconomic characteristics by region

The complete set of ECHO database includes all public hospital discharges produced in the ECHO countries (Denmark, England, Portugal, Slovenia and Spain) from 2002 to 2009. This represents a total of 191.136.051 individual discharges assigned to a specific hospital and to the region of residence of the patient.

4.1. Sources and data availability

Table 2 identifies the original sources of data that were used to construct the ECHO database. All countries provided data from 2002 to 2009 except Austria. This data is available in the ECHO database with the exception of Slovenian data between 2002 and 2004 due to missed data in diagnoses and procedures (ECHO, 2014).

Table 2: Sources of data

Country	**Inpatient Hospital Episodes**	**Hospital Resources**	**Regional Profile**
Denmark	*Ministeriet Sunhed Forebyggelse*	*Ministeriet Sunhed Forebyggelse*	*Danmarks Statistik*
England	NHS England	Department of Health	UK National Statistics
Portugal	*Ministério da Saúde*	*Ministério da Saúde*	*Instituto Nacional de Estatística (INE) Portugal*
Slovenia	*Ministrstavo za Zdravje*	*Ministrstavo za Zdravje*	*Statistični Urad Republike Slovenije*
Spain	ATLAS *de Variaciones en la Práctica Médica* (ATLAS-VPM)	*Ministerio de Sanidad, Servicios Sociales e Igualdad*	*Instituto Nacional de Estadística (INE) Spain*

Source: ECHO (2014)

4.2. Inpatient hospital episodes

Patient-level data on hospital admissions provides demographic and clinical information on each inpatient.

Table 3: Inpatient hospital episodes data details

Variable	Description	Type
	Hospital Identification	
Hospital code	Identification of the hospital where the patient was admitted.	Nominal
	Demographic information	
Gender	Identification of gender of patient: 0 if man and 1 if women.	Nominal
Age at admission	Identification of age of patient. Birthday date is also available.	Continuous
Zip-code	Identification of residence of the patient by the Smallest Unit of Analysis available.	Nominal
	Clinical information	
Admission/Discharge date	Date when patient was admitted in the hospital and date when patient was discharged from the hospital.	Date
Length of stay	Number of days the patient spent at the hospital.	Continuous
Discharge status of patient	Reasoning behind discharge: ▪ Patient went home ▪ Transferred to another hospital unit ▪ Death ▪ Left against medical order ▪ Other	Nominal
Diagnostics	Identification of patient's diagnostics: ▪ Admission diagnostic – diagnostic that generated the admission; ▪ Primary diagnostic – the principal condition; it can be different from the diagnostic that generated the admission; ▪ Secondary diagnostics – additional conditions; Disease classification systems differ between countries (ECHO, 2014; Hindle, 2003; Kobel *et al.*, 2011): ▪ Denmark – International Classification of Diseases, 10th version (ICD-10) ▪ England – ICD-10	Nominal

Variable	Description	Type
	▪ Portugal – International Classification of Diseases, 9th version, Clinical Modification (ICD-9-CM) ▪ Slovenia – International Classification of Diseases, 10th version, Australian Modification (ICD-10-AM) ▪ Spain – ICD-9-CM	
Procedures	Identification of procedures performed during inpatient stay. Procedure classification systems differ between countries (ECHO, 2014; Hindle, 2003; Kobel *et al.*, 2011): ▪ Denmark – Nordic Medico-Statistical Committee (NOMESCO) Classification of Surgical Procedures (NCSP) ▪ England – Office of Population Censuses and Surveys (OPCS) Classification of Interventions and Procedures ▪ Portugal – ICD-9-CM ▪ Slovenia – Australian Classification for Health Intervention (ACHI) ▪ Spain – ICD-9-CM	Nominal
Patient classification	Classification of patient according to their diagnostics and procedures underwent during in-patient stay. Countries in the database use different patient classification systems (Hindle, 2003; Kobel *et al.*, 2011): ▪ Denmark – Danish patient classification system (Dk NordDRG) ▪ England – Healthcare Resource Group (HRG) ▪ Portugal – All-Patients Diagnostic Related Groups (AP-DRG) ▪ Slovenia – Australian-Refined Diagnostic Related Groups (AR-DRG) ▪ Spain – AP-DRG	Nominal

4.3. Hospital resources

Hospital resources data provides information on the resources available at each hospital, i.e., resources that the hospital has available to produce healthcare services and to treat patients.

Table 4: Hospital resources data details

Variable	Description	Type
Human resources		
Employees	Total number of human resources that work in the hospital. It counts not only medical staff but also administrative staff.	Discrete
Doctors	Total number of doctors that work in the hospital regardless of their specialty.	Discrete
Surgical specialties doctors	Total number of doctors from surgical specialties, working in the hospital.	Discrete
Medical specialties doctors	Total number of doctors from medical specialties, working in the hospital.	Discrete
Orthopedic surgeons	Total number of orthopedic surgeons available at the hospital.	Discrete
Gynaecology/obstetric doctors	Total number of gynecology/obstetric doctors available at the hospital.	Discrete
Nurses	Total number of nurses available at the hospital.	Discrete
Interns	Total number of interns doing their internship at the hospital. [This information can also be grouped by Teaching status of the hospital – whether the hospital is a Teaching hospital or not.]	Discrete [Binary]
Physical resources		
Beds	Total number of beds, regardless of the unit, available at the hospital.	Discrete
ICU beds	Total number of beds allocated to intensive care unit.	Discrete

Variable	Description	Type
Orthopedic beds	Total number of beds allocated to orthopedic units.	Discrete
Surgical beds	Total number of beds allocated to surgical procedures.	Discrete
Neonatal ICU beds	Total number of beds allocated to neonatal intensive care unit.	Discrete
Obstetric/ gynecologic beds	Total number of beds allocated to obstetric/gynecologic unit.	Discrete
Specialties	Number of different specialties available at the hospital.	Discrete
Hemodynamic Units	Dummy variable stating whether the hospital has a hemodynamic unit.	Binary
Linear accelerators	Number of linear accelerators available at the hospital.	Discrete

4.4. Regional profile

This set of data refers to socioeconomic characteristics of regions that may affect demand for healthcare services. Table 5 presents the Smallest Unit of Analysis available in the database for each country (Bernal-Delgado *et al.*, 2015).

Table 5: Smallest Units of Analysis in ECHO database

	Denmark	Portugal	Slovenia	Spain	UK
Smallest Unit of Analysis	Kommuner	Concelhos	Statistical Region	Healthcare Areas	Health Authorities
#	98	278	12	199	326

Source: Bernal-Delgado et al. (2015)

Table 6: Regional data details

Variable	Description	Type
Regional profile		
Population per smallest unit of analysis	Number of inhabitants per region by age group.	Discrete
Average family income per capita	Average Family income divided by the average number of individuals in the family.	Continuous
Education	Population achieving a specific level of education:	
	▪ Proportion of population attaining primary level as the highest completed level of education;	Continuous
	▪ Proportion of population attaining secondary level as the highest completed level of education;	Continuous
	▪ Proportion of population attaining university level as the highest completed level of education;	Continuous
Medical specialties doctors	Total number of doctors from medical specialties, working in the hospital.	Discrete
Unemployment	Population unemployed:	
	▪ Number of active people looking for a job;	Discrete
	▪ Proportion of active population looking for a job;	Continuous
	▪ Proportion of active men looking for a job;	Continuous
	▪ Proportion of active women looking for a job;	Continuous
	▪ Proportion unemployment by age groups;	Continuous
Urbanization Degree	Classification of degree of urbanization of regions (Conselho Superior de Estatística, 2009):	Ordinal
	▪ Urban: Pop/Km^2 > 500;	
	▪ Semi-urban: 100< Pop/Km^2 < 500;	
	▪ Rural: Pop/Km^2 < 100.	

4.5. Structure of data

Figure 7 presents a scheme on how the data is structured and how the different categories are linked.

Figure 7: Scheme of database

In-patient hospital episodes

Inpatient	Hospital Code	Gender	…	Zip Code	Primary Diagnostic	…
1	P101	0		PT111	430.56	
2	P078	0		PT11B	758.01	
3	P003	1		PT186	341.29	

Hospital resources

Hospital Code	Beds	Employees	…
P001	236	1.305	
P002	132	789	
P003	76	561	

Socioeconomic characteristics by region

Zip Code	Population	Average Family Income	…
PT111	36.759	948,45€	
PT112	67.198	1.136,98€	
PT119	123.567	1.693,56€	

For the different objectives, subsets of data were extracted from this database. Detailed information on the data used for each objective is available in the presentation of methods of each work produced.

5. Results

Portugal: Geographic variations in health care (Work 1)

Céu Mateus
Inês Joaquim
Carla Nunes
Paulo Boto
Luís Campos

in *Geographic Variations in Health Care: What Do We Know and What Can Be Done to Improve Health System Performance?*

OECD Publishing, Paris

2014

DOI: http://dx.doi.org/10.1787/9789264216594-14-en

Chapter 11

Portugal: Geographic variations in health care

Céu Mateus, Inês Joaquim and Carla Nunes,
National School of Public Health, Nova University of Lisbon,

Paulo Boto, National School of Public Health, Nova University of Lisbon,
Ministry of Health,

and

Luís Campos, Centro Hospitalar Lisboa Ocidental, Lisbon,
and Faculty of Medical Sciences, Nova University of Lisbon

During the eight-year period covered in this report (2002-09), there has been a reduction in geographic variations in the use of some of the health care procedures selected in this study, such as cardiac catheterisation and coronary angioplasty. There has also been a reduction in geographic variations in caesarean section rates, although the variation between public and private hospitals rates continues to be wide. The Portuguese Ministry of Health recently asked a group of experts to develop a plan to reduce the inappropriate use of caesarean sections throughout the country.

This study also shows that geographic variations in the use of some other procedures that are becoming less frequently used and replaced by other treatment options are increasing, for example for coronary artery bypass graft (CABG) and hysterectomy. This indicates that the reduction in the use of these procedures has not been uniform across the countries, and a need to promote greater convergence in clinical practices.

11.1. Introduction

Life expectancy for both men and women in Portugal has continued to increase between 2002 and 2012, while the crude and standardised death rates have decreased. However, regional disparities persist, particularly between urban-coastal regions and rural-interior regions, with worst health and living conditions in the latter regions (Barros et al., 2011).

Up until recently, there had been little, if any, studies of medical practice variations in Portugal. In 2010, a group of researchers based at the National School of Public Health became involved in the ECHO (European Collaboration for Healthcare Optimization) project, funded by the European Commission. The project results will be the first to analyse medical practice variations at a geographic and hospital level in Portugal. The findings presented in this chapter draw on the ECHO work.

Section 11.2 provides an overview of the Portuguese health care system. Section 11.3 turns to the data and methods used, followed by a presentation of the results. Between 2002 and 2009, there has been a slight reduction in the geographic variations in the use of most of the procedures covered under this study. However, in 2009, there were still substantial variations in the utilisation rate of selected procedures in Portugal which cannot be solely attributed to differences in population structure. The final section provides some conclusions and policy discussions. It is important to address medical practice variations, because equitable access to health care is a cherished goal of Portuguese health care policy and of the NHS in particular. Identifying areas of both appropriate and inappropriate care will help to understand the determinants of good performance and create opportunities to monitor the impact of changes.

11.2. Overview of Portugal's health care system

Political and organisational structure

Portugal has a tax-funded national health service that provides coverage to all residents. Health care is a shared responsibility between the central and regional level. The Ministry of Health and its institutions oversee the planning and regulation of the national health service (NHS). The Ministry of Health is responsible for the design, implementation and evaluation of the overall health plan, while the five regional health administrative boards look after the management and delivery of health services. The health administrative boards are accountable to the Ministry of Health and define the strategic management of the population's health, the supervision and control of hospitals, the management of primary care and NHS primary care centres, as well as the contractual agreements for services with hospitals and private sector providers (Barros et al., 2011).

Health care expenditure

Total health spending accounted for 10.2% of GDP in Portugal in 2011, above the OECD average of 9.3% (OECD, 2013a). However, Portugal ranks below the OECD average in terms of health spending per capita, with spending of USD 2 619 in 2011 (adjusted for purchasing power parity), compared with the OECD average of around USD 3 300. The share of hospital spending in Portugal in 2011 (27%) was slightly lower than the OECD average (29%).

Health spending per capita in Portugal increased in real terms by 1.8% on average between 2000 and 2009, and grew at a similar rate of 2% for 2009/10. However, health spending per capita fell by 6.3% in 2010/11. Several other OECD countries also experienced a marked slowdown or reduction in health spending after 2010, following the recession and the need for fiscal consolidation.

Health care financing

The NHS provides universal health coverage for all residents, and health care is largely financed from general taxation. Public spending accounted for 65% of total health spending in 2011, less than the OECD average of 72% (OECD, 2013a). About one-fifth of the population also have a private health insurance (PHI), which provides duplicate coverage (i.e., faster access to health services in the private sector). PHI accounted for only 5% of total health spending in 2011, with the remainder (about 30%) financed from direct out-of-pocket payments by households (OECD, 2013a).

Health care delivery and provider payments

Physician services and payments

Primary care in the NHS is predominantly delivered in public clinics staffed by physicians and other health professionals (OECD Health Systems Characteristics Survey, 2012b). Physicians in the NHS play a gatekeeper role in primary care and are public salaried employees, though those working in family health units partially receive remuneration that includes capitation (risk-adjusted) and pay-for-performance (OECD Health Systems Characteristics Survey, 2012b). Specialists provide services in inpatient and outpatient departments of hospitals, and are also public salaried employees. Private sector providers are remunerated mainly on a fee-for-services basis (Barros et al., 2011).

Hospital services and payments

Hospital services are provided by both the public and the private sector. NHS hospitals provide elective and non-elective care, ambulatory surgery, maternity services, diagnostic procedures, ancillary tests, and accident and emergency services. Most non-acute psychiatric inpatient and outpatient services are provided by psychiatric hospitals.

Public hospitals (which accounted for 72% of all hospital beds in 2011) are funded through prospective global budgets, with the financing of inpatient and ambulatory surgery based on diagnosis related groups (DRGs). Since 2012, 4% of the budget is allocated based on quality measure improvements (OECD Health Systems Characteristics Survey, 2012b). Private not-for-profit hospitals (20% of beds) and private for-profit hospitals (8%) are remunerated on a fee-for-service basis (OECD, 2013b).

In terms of activities, 80% of inpatients were cared for in NHS public hospitals in 2010, with the remaining 20% treated in private hospitals (DGS, 2012). Hospital activities generally increased over the past decade, in particular in the areas of ambulatory surgeries, day hospital sessions and consultations.

Several important measures have been taken in recent years to reshape the provision of health care in Portuguese NHS hospitals. One that is particularly noteworthy concerns administrative hospital mergers. From 2000 to 2010, more than 50 hospitals were merged with others, although no hospital was closed. Other policy measures include support for the implementation of public-private partnerships and the creation of Local Health Units, which combine under the same board the management of hospitals and primary care centres (Barros et al., 2011).

11.3. Data and methods

Data for the selected procedures were obtained from the national DRGs database, which includes all inpatient episodes in the Portuguese NHS hospitals in any given year. In Portugal, coding is performed by physicians with specific training. During the period under analysis, two different groupers were used in Portuguese NHS hospitals: HCFA DRG version 16.0 until mid-2006, and AP-DRG version 21.0 from mid-2006 onwards. Diagnoses and procedures were coded based on the ICD-9-CM classification.

The list of procedures selected is displayed in the results presented in Table 11.1. Surgery after hip fracture is expected to be a low variation procedure and was used for calibration purposes.

Table 11.1. Total inpatient discharges and ALOS, and discharges for selected procedures, Portugal, 2002-09

	2002	2003	2004	2005	2006	2007	2008	2009	% change 2002-09
Inpatient discharges	957 592	977 523	971 597	983 004	975 185	964 784	959 036	936 315	-2.20%
ALOS[1] (days)	7.30	7.02	7.05	7.05	7.02	7.10	7.11	7.00	-4.10%
Medical admissions	465 615	465 316	462 767	473 691	459 352	457 412	456 233	454 750	-2%
CABG	2 379	2 283	2 236	2 355	2 588	2 556	2 500	2 467	4%
PTCA	3 982	5 667	6 737	7 614	8 190	9 505	9 914	9 715	144%
Cardiac catheterisation	12 121	13 693	15 589	16 074	17 900	24 438	24 814	24 224	100%
Surgery after hip fracture	8 476	9 035	9 197	9 178	9 001	9 552	9 921	10 344	22%
Knee replacement	2 764	3 417	4 196	4 359	5 308	5 384	6 091	6 601	139%
Knee arthroscopy	3 518	3 853	4 663	4 695	5 023	5 516	5 392	5 924	68%
Caesarean sections	27 917	28 332	28 060	28 985	28 238	27 605	28 181	26 859	-4%
Hysterectomy	11 445	12 049	12 007	11 693	11 545	11 087	10 288	7 290	-36%

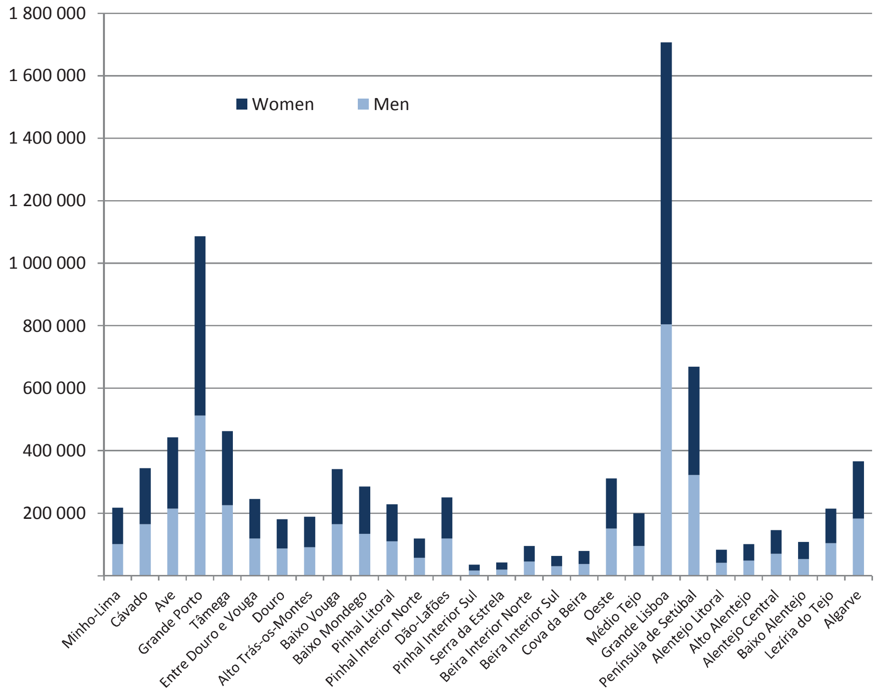
1. ALOS: Average length of stay. Inpatient only.

Source: Authors' estimates based on national DRG database, 2002-2009.

The total number of inpatient discharges in NHS hospitals decreased by 2.2% from 2002 to 2009, and the average length of stay (ALOS) also decreased by 4.1%. The number of medical admissions (which represents slightly less than half of all inpatient admissions), caesarean sections and hysterectomies also decreased during this period, although the rate of caesarean sections per 1 000 live births increased from 30% to 36% between 2002 and 2009. The number of cardiac catheterisation (used to diagnose ischaemic heart diseases) doubled, with the number of coronary artery bypass graft (CABG) increasing by a modest 4% (and declining in fact in the most recent years) while coronary angioplasty (PTCA) grew steadily and rapidly over this period. As for knee procedures, the number of knee replacement more than doubled (from 2 764 to 6 601), while the number of knee arthroscopies rose by 68%. Finally, the number of surgery after hip fracture increased by 22%.

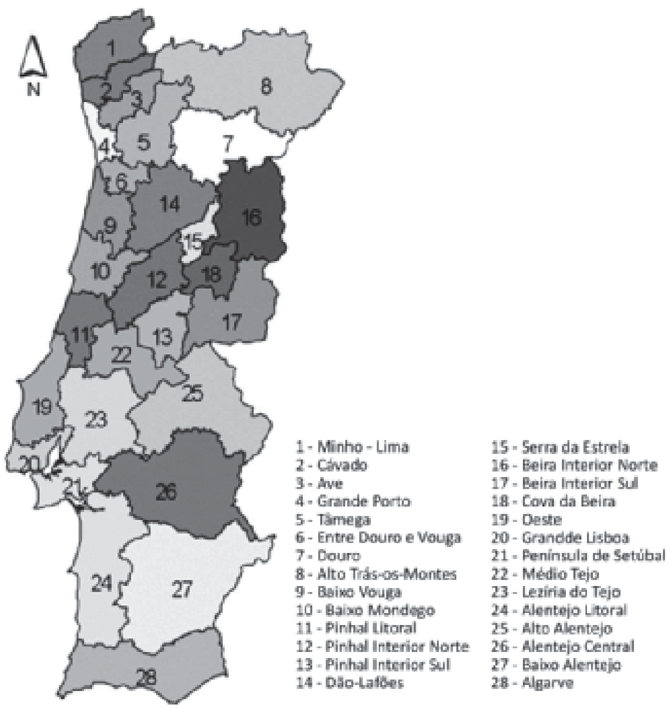
The geographic unit of analysis in this report is based on NUTS III, which corresponds to 28 groups of municipalities in Portugal inland. Figure 11.1 shows the size of the population and the gender breakdown in these 28 groups of municipalities and Figure 11.2 displays a map of the Portuguese municipalities. Standardised rates were based on the national population structure of 2009 (INE, 2013).

Figure 11.1. Population over 15 by gender and geographic unit, Portugal inland, 2009



Source: National Statistical Institute (INE), annual estimates of the resident population in Portugal, 2009.

Figure 11.2. Map of the Portuguese municipalities



Source: http://pt.wikipedia.org/wiki/Ficheiro:NUTS_III.png#file.

Table 11.2 provides some key facts on the geographic disparities in the supply of hospital beds and different categories of doctors across these 28 regions. The number of beds in NHS hospitals ranges from 72 per 100 000 population in Serra da Estrela to 878 in Baixo Mondego. The region of Baixo Mondego also has the highest rates for the different categories of doctors shown in this table. The number of cardiac surgeons per 100 000 population is higher in those regions that have teaching hospitals: Baixo Mondego, Grande Lisboa and Grande Porto. It should be noted that only cardiac surgeons do CABG procedures, while PTCAs and catheterisations are performed by cardiologists. In small regions that are sparsely populated such as Serra da Estrela, there are no medical specialists.

Table 11.2. Beds, gynaecologists-obstetricians, orthopaedic surgeons and cardiac surgeons per 100 000 population, by geographic region, Portugal, 2010

	Beds	Gyn-obstetrics physicians	Orthopaedic surgeons	Cardiac surgeons
Minho-Lima	187	6	4	0.4
Cávado	164	5.3	5.8	3.4
Ave	164	8.2	6.7	3.4
Grande Porto	278	11.4	7.4	8.2
Tâmega	90	3.2	3	1.6
Entre Douro e Vouga	141	5.9	5.2	2.4
Douro	251	4.3	7.2	4.3
Alto Trás-os-Montes	321	3.8	6.6	1.4
Baixo Vouga	144	3.7	5	2.7
Baixo Mondego	878	31.9	17.3	18.3
Pinhal Litoral	209	6.7	5.6	2.2
Pinhal Interior Norte	Not available			
Dão-Lafões	222	6.5	7.9	3.4
Pinhal Interior Sul	Not available			
Serra da Estrela	72	0	0	0
Beira Interior Norte	303	7.4	7.4	4.6
Beira Interior Sul	420	4.1	5.5	6.9
Cova da Beira	378	18.9	5.6	4.4
Oeste	127	4.4	4.1	0.8
Médio Tejo	205	3.9	6.1	2.6
Grande Lisboa	292	8.6	5.2	9.3
Península de Setúbal	178	6.3	4.5	3.8
Alentejo Litoral	121	0	3.2	0
Alto Alentejo	255	2.6	5.2	0.9
Alentejo Central	197	4.8	4.2	4.2
Baixo Alentejo	186	6.4	4	1.6
Lezíria do Tejo	153	6	3.6	3.6
Algarve	188	6	4.6	2.8
National average	239	7.8	5.8	5.3

Source: National Statistical Institute (INE), Hospitals' Survey, 2010.

11.4. Description of results

Overview of results

In 2009, geographic variations among the selected set of health care activities and procedures in Portugal was highest for CABG, and for knee replacement and knee arthroscopy (Table 11.3). It was the lowest for caesarean sections and surgery after hip fracture. The low variation for surgery after hip fracture was expected, given that this was selected as a “calibration” procedure on the grounds that there is little discretion for doctors to operate patients suffering from a hip fracture. Regarding caesarean sections, while the degree of variations across different regions in Portugal is low, the rates have generally increased in most regions between 2002 and 2009, and were much higher in 2009 than in most regions in Spain (see the chapter on Spain in this publication).

Table 11.3. Summary of geographic variations for a selected set of health care activities and procedures, Portugal, 2009

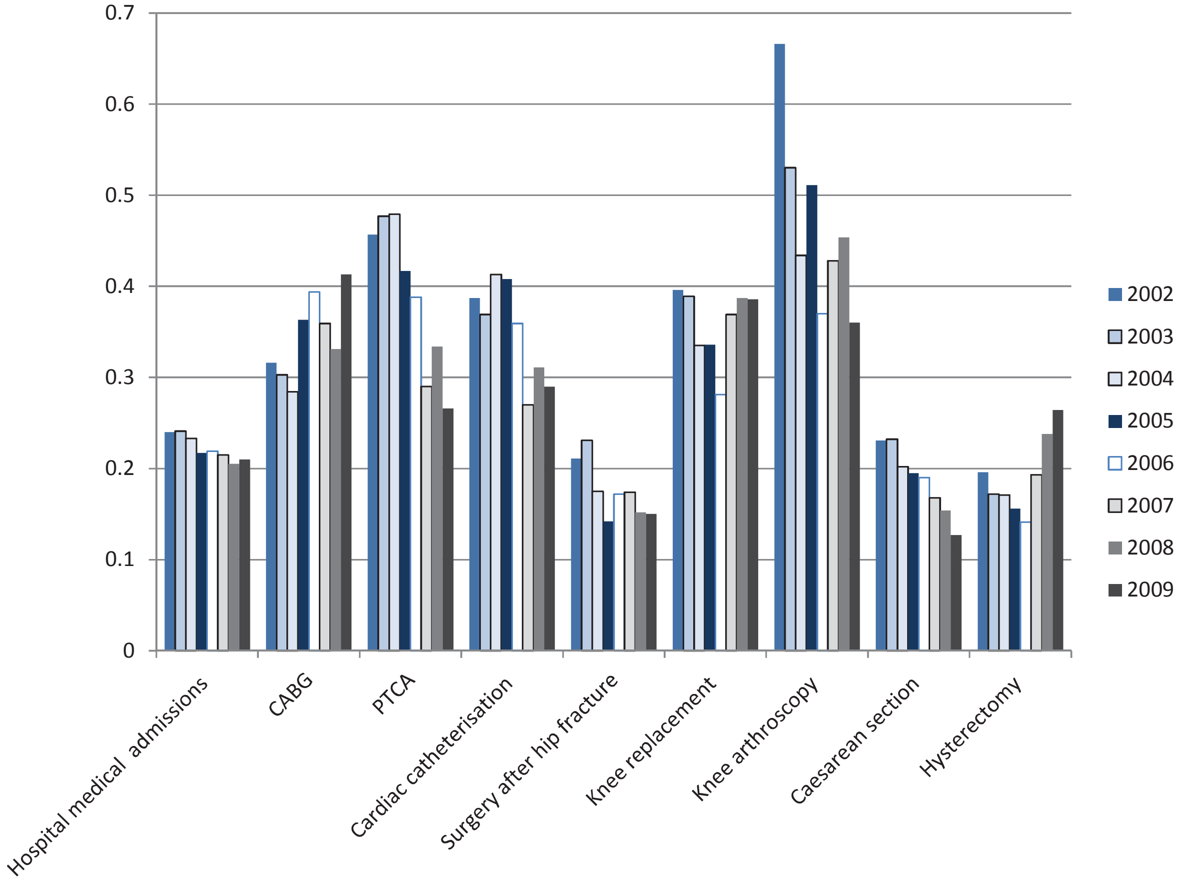
	Hospital medical admissions	CABG	PTCA	Catheterisation	Surgery after hip fracture	Knee replacement	Knee arthroscopy	Caesarean section (per 1 000 live births)	Hysterectomy
Crude rate (national)	4 483	24	96	239	102	65	58	329	175
Standardised unweighted average rate	5 569	27	117	327	126	84	67	330	179
Q10 (lowest decile)	4 449	16	83	235	107	42	42	278	121
Q90 (highest decile)	6 462	41	154	472	147	123	106	397	238
Coefficient of variation	0.21	0.41	0.27	0.29	0.15	0.39	0.36	0.13	0.26
Systematic component of variation	5	13.6	6.4	10.3	2.3	18	11.3	1.3	7.5

Note: All rates are expressed per 100 000 population, except caesarean sections (per 1 000 live births) and hysterectomy (per 100 000 women).

Source: Authors’ estimates based on national DRG database, 2002-2009.

Figure 11.3 presents trends over time in the coefficient of variation for the selected health care activities and procedures. While there has been an increase in regional variations in the use of CABG between 2002 and 2009, this has been accompanied by a reduction in variations in PTCA rates, indicating that there was some convergence in the use of PTCA across regions. Regional variations in knee arthroscopy decreased, but still remain very high. Following some reductions in regional variations for knee replacement between 2002 and 2006, the degree of variations went up again between 2007 and 2009, so there was no reduction over the entire period. While the overall number of hysterectomies in Portugal has come down significantly between 2002 and 2009 (Table 11.1), this has been accompanied by a rise in regional variations in hysterectomy rates, indicating that the reduction has not been uniform across the country.

Figure 11.3. Evolution of the coefficient of variation across geographic regions for a selected set of health care activities and procedures, Portugal, 2002-09

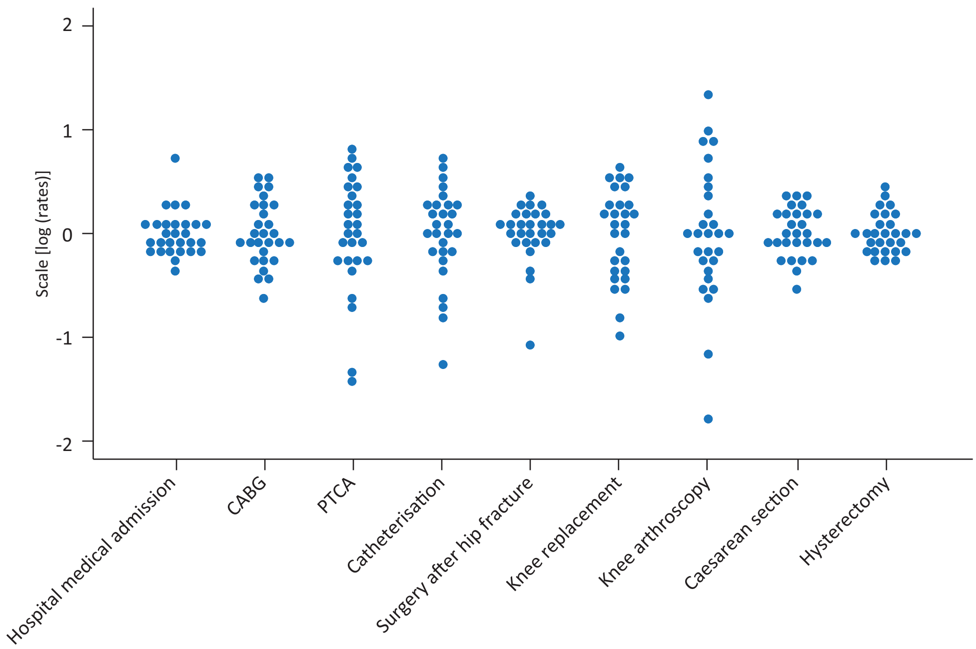


Source: Authors' estimates based on national DRG database, 2002-2009.

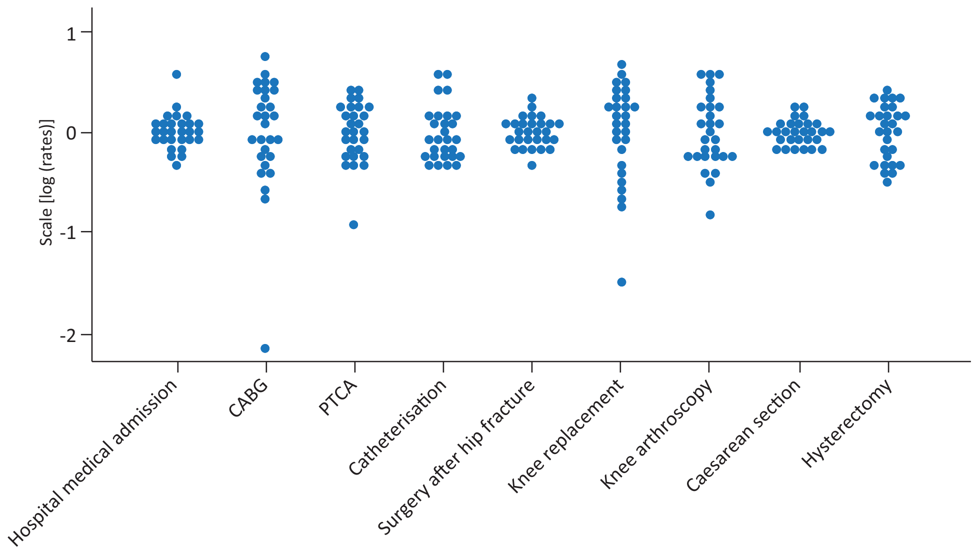
Figure 11.4 shows the regional variations in the log of the standardised rate for each procedure, with the values centered on the national average for each procedure, in 2002 and 2009. It illustrates in another way that the variations are more marked for cardiac care and knee procedures. In general, there is always more dispersion for the regions below the zero line, meaning that those with rates below the national average are more different from the national pattern than those above the average.

Figure 11.4. Geographic variations for a selected set of health care activities and procedures, Portugal, 2002 and 2009

2002



2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Hospital medical admissions

The level of geographic variations in hospital medical admissions in Portugal was relatively low in 2009, and declined slightly between 2002 and 2009, in a context of a slight overall reduction in hospital medical admission rates during that period (Table 11.4).

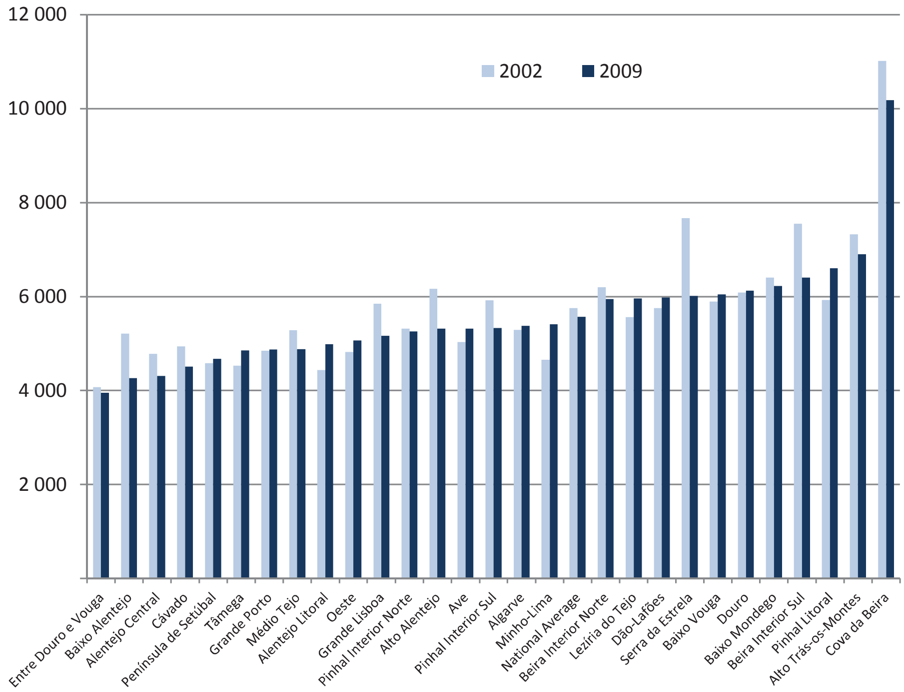
Table 11.4. Hospital medical admissions standardised rate per 100 000 population, Portugal, 2002-09

NUTS III	2002	2003	2004	2005	2006	2007	2008	2009
Unweighted average	5 755	5 774	5 771	5 842	5 702	5 593	5 586	5 569
Q10	4 566	4 683	4 639	4 682	4 651	4 507	4 490	4 449
Q90	7 393	7 625	7 332	7 452	7 241	6 843	6 499	6 462
Coefficient of variation	0.24	0.24	0.23	0.22	0.22	0.22	0.21	0.21

Source: Authors' estimates based on national DRG database, 2002-2009.

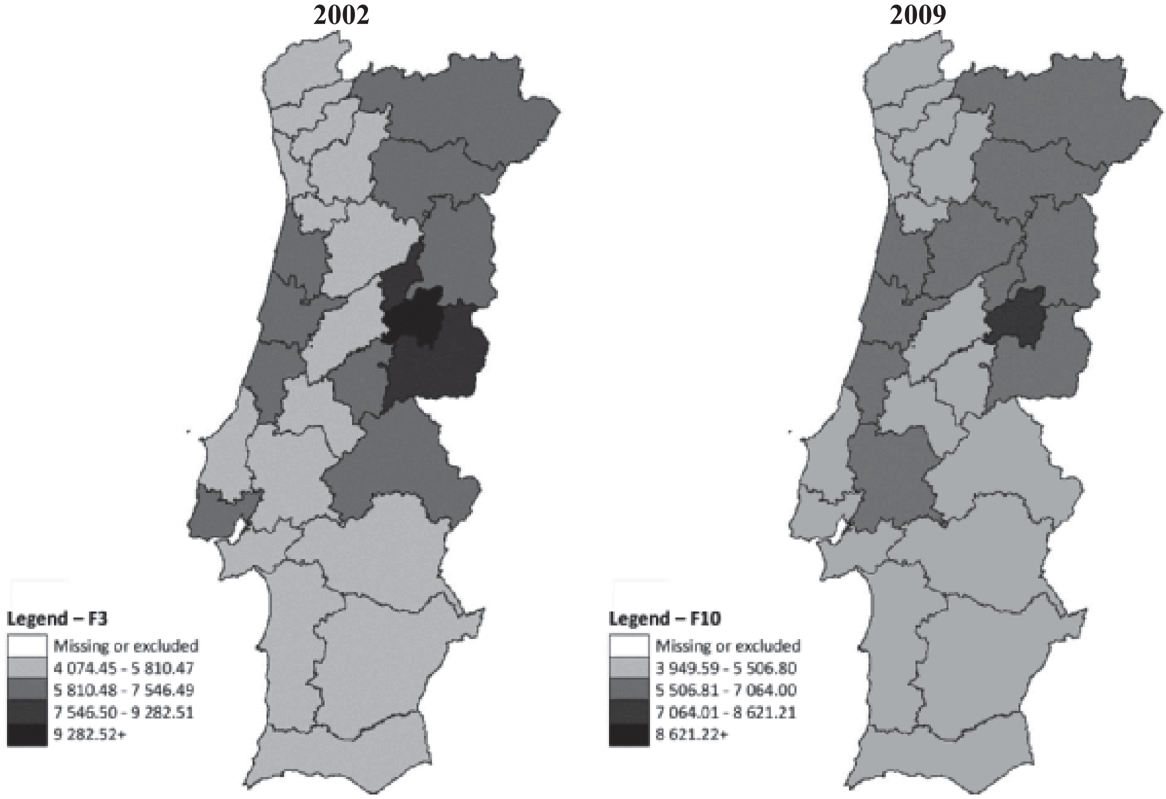
In 2009, Cova da Beira was the region with the highest rate of hospital medical admissions (standardised for age and sex), despite the fact that admission rates decreased by about 10% between 2002 and 2009 (Figure 11.5). Following Cova da Beira, the regions of Alto Trás-os-Montes and Pinhal Litoral had the highest rates in 2009, with all three regions having rates at least 50% higher than in the Entre Douro e Vouga and Baixo Alentejo regions. In general, the rural-interior regions located in the northeast part of the country tend to have higher hospital medical admission rates than the urban-coastal regions (Figure 11.6).

Figure 11.5. Hospital medical admissions standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.6. Maps of hospital medical admissions standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Revascularisation procedures

Cardiovascular diseases, the leading cause of mortality in Portugal, are responsible for around 30% of all deaths. The growing number of cardiac catheterisation and revascularisation procedures to treat people with ischaemic heart disease certainly contributed to the reduction in (age-standardised) mortality rates from cardiac disease between 2002 and 2009.

CABG

While CABG rates remained fairly stable overall in Portugal between 2002 and 2009, the coefficient of variation across regions increased during this period, although there were fluctuations from year-to-year (Table 11.5 and Figure 11.7).

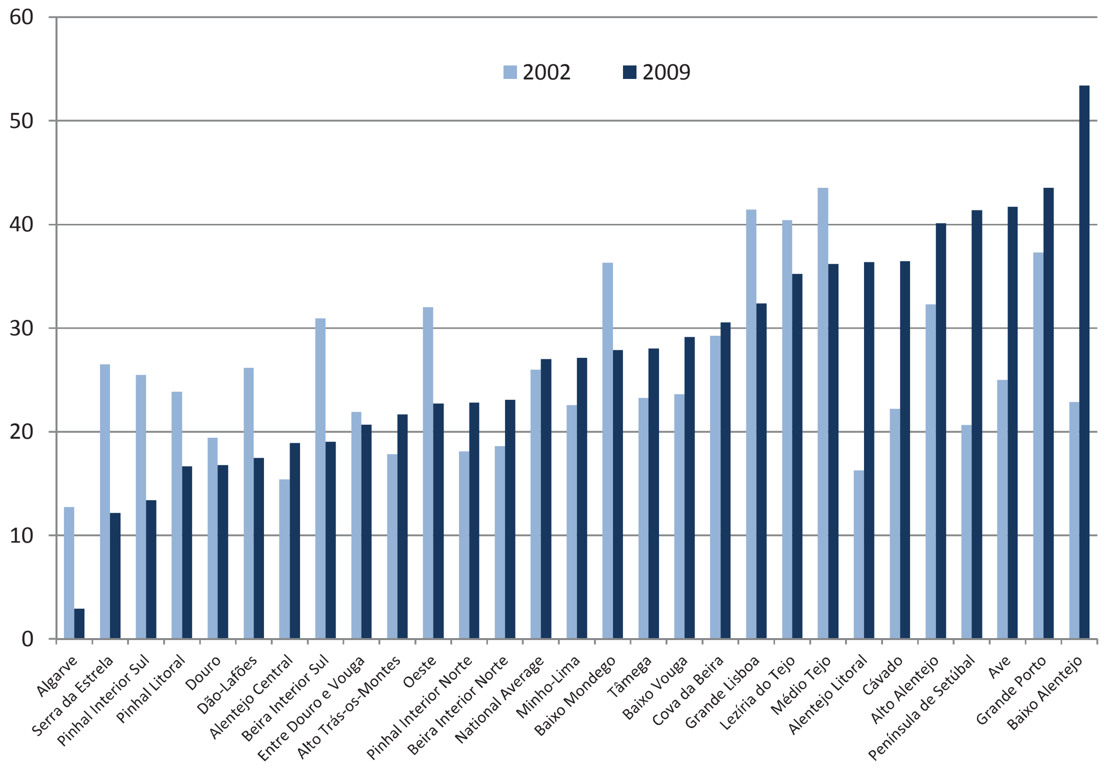
Table 11.5. Coronary artery bypass graft (CABG) standardised rate per 100 000 population, Portugal, 2002-09

NUTS III	2002	2003	2004	2005	2006	2007	2008	2009
Unweighted average	26	25	25	26	26	28	28	27
Q10	17	15	16	13	13	18	18	16
Q90	38	37	35	37	41	40	41	41
Coefficient of variation	0.32	0.3	0.28	0.36	0.39	0.36	0.33	0.41

Source: Authors' estimates based on national DRG database, 2002-2009.

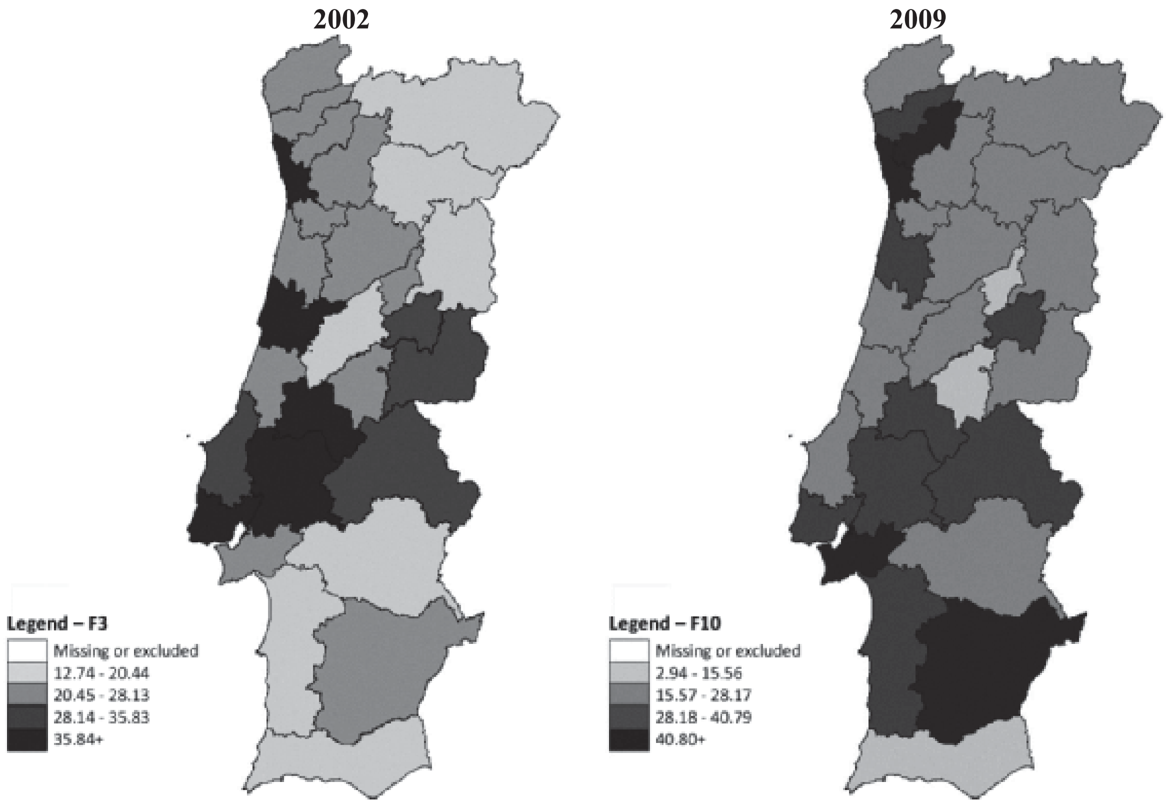
The use of CABG (a very invasive procedure, involving an open chest surgery) has decreased in most regions, being replaced by coronary angioplasty or other less invasive treatments for ischaemic heart diseases, but CABG rates have also increased in some regions, like Baixo Alentejo, Península de Setúbal, Alentejo Litoral and Ave (Figure 11.8). This might possibly indicate growing variations in the treatment of ischaemic heart diseases, with a possible over-use of CABG in some regions, although this might also reflect a growing concentration of CABG surgery in certain regions and hospitals.

Figure 11.7. CABG standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.8. Maps of CABG standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

PTCA

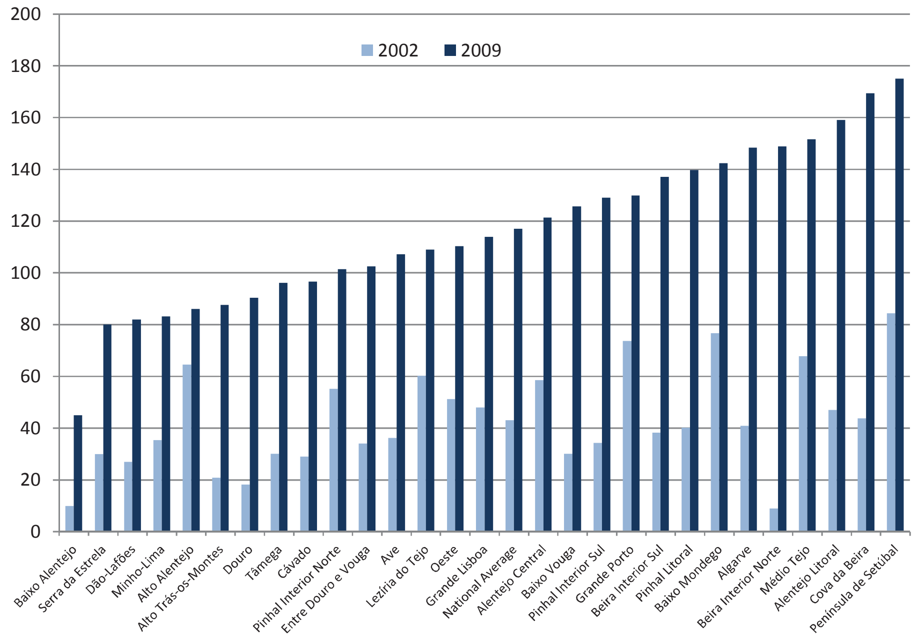
For PTCA, the unweighted average rate across regions nearly tripled over the period (Table 11.6). Combined with the fact that the coefficient of variation fell substantially during this period, this means that the growth rate was particularly strong in those regions that had relatively low rates in 2002, pointing towards some convergence in the use of PTCA across regions (Figures 11.9 and 11.10). The growth rate of PTCA was marked in the regions of Beira Interior Norte, Douro, Baixo Alentejo, Alto Trás-os-Montes, and Baixo Vouga.

Table 11.6. Coronary angioplasty (PTCA) standardised rate per 100 000 population, Portugal, 2002-09

NUTS III	2002	2003	2004	2005	2006	2007	2008	2009
Unweighted average	43	55	68	79	89	108	121	117
Q10	20	26	31	39	50	69	86	83
Q90	70	86	113	118	138	153	166	154
Coefficient of variation	0.46	0.48	0.48	0.42	0.39	0.29	0.33	0.27

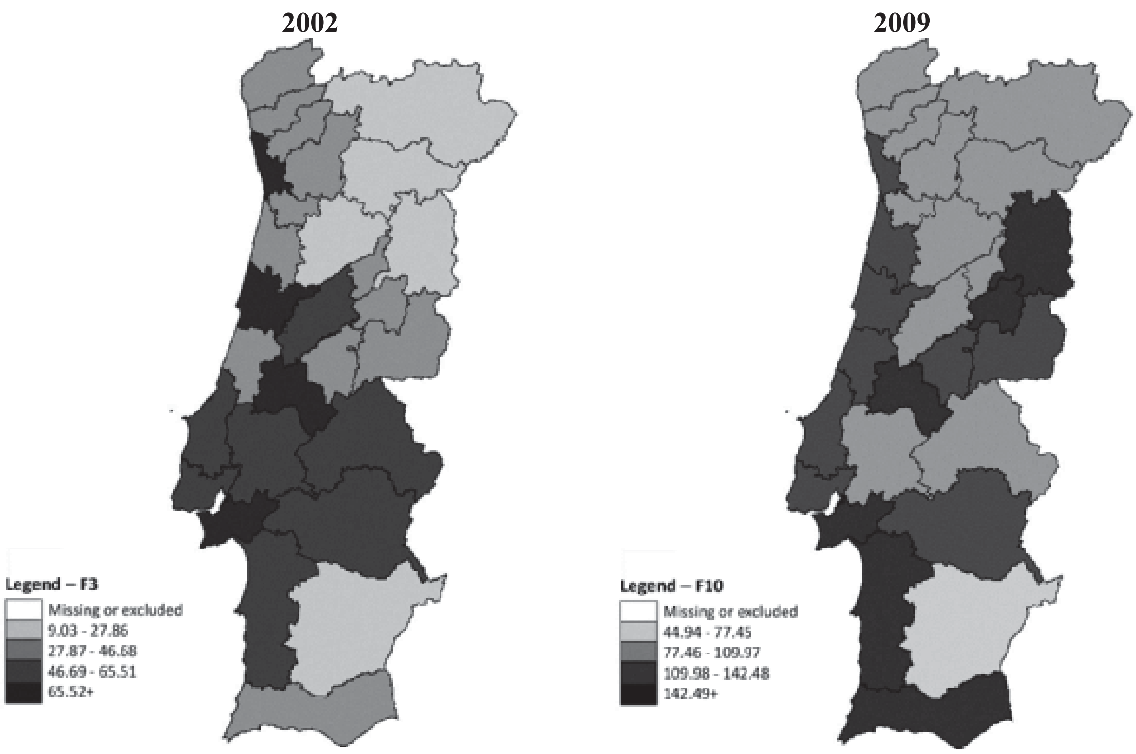
Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.9. PTCA standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.10. Maps of PTCA standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Catheterisation

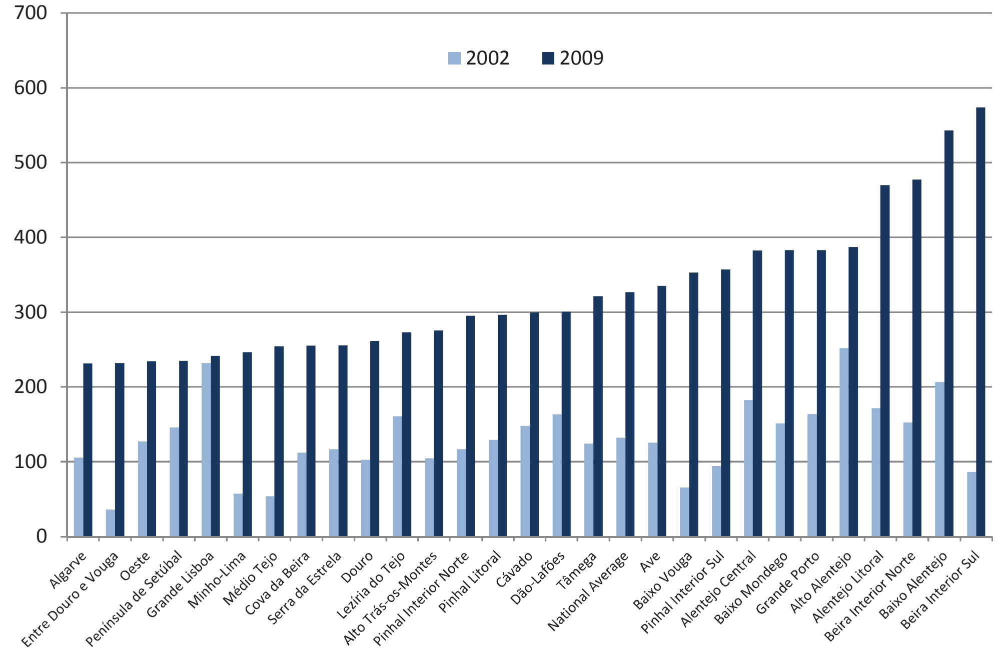
The trend rise in cardiac catheterisation (used to diagnose ischaemic heart disease) is, not surprisingly, fairly similar to the rise in PTCA. The unweighted average rate of cardiac catheterisation more than doubled between 2002 and 2009, while the coefficient of variation went down, indicating a more uniform access to this important diagnostic procedure (Table 11.7, Figures 11.11 and 11.12). The Grande Lisboa is an exception to the general strong growth in cardiac catheterisation rates: the rate in the national capital region was one of the highest in 2002, but did not increase much up in the following years, with the result that the region had one of the lowest rates in 2009, well below the national average.

Table 11.7. Cardiac catheterisation standardised rate per 100 000 population, Portugal, 2002-09

NUTS III	2002	2003	2004	2005	2006	2007	2008	2009
Unweighted average	132	156	176	180	214	323	341	327
Q10	63	85	74	78	117	239	232	235
Q90	190	219	267	267	290	418	489	472
Coefficient of variation	0.39	0.37	0.41	0.41	0.36	0.27	0.31	0.29

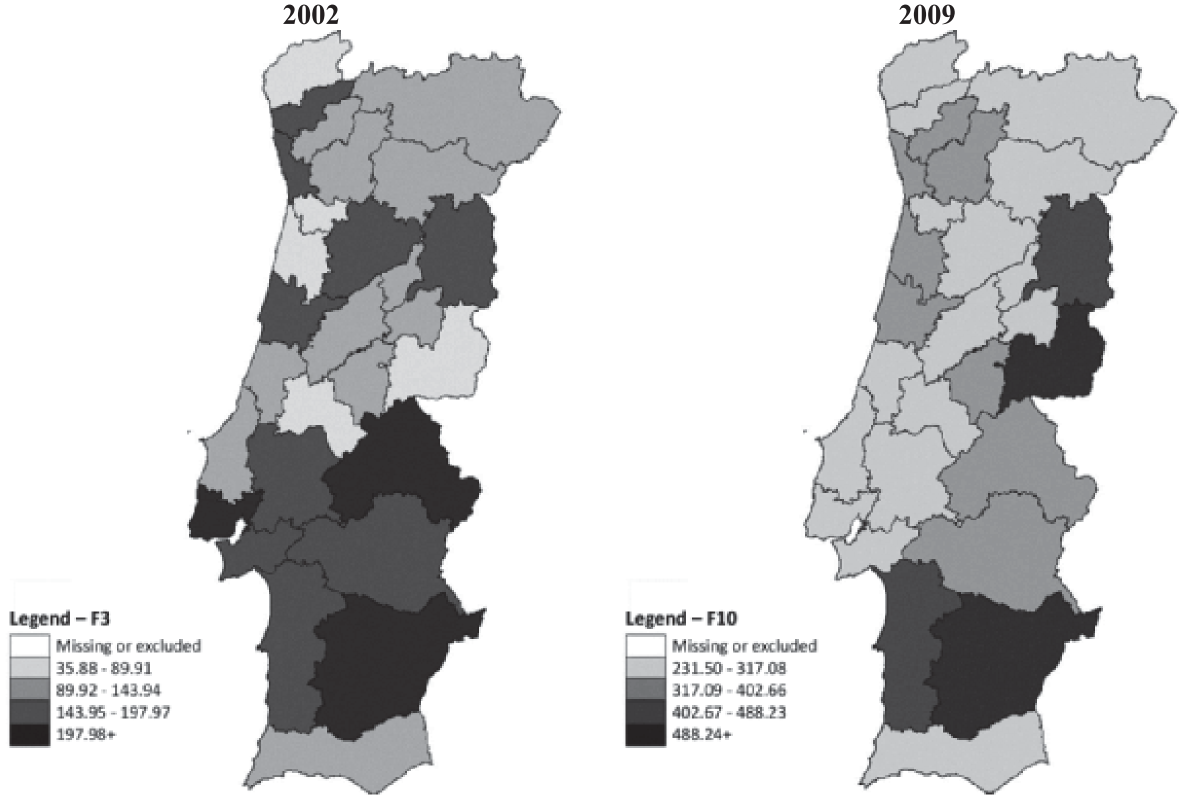
Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.11. Cardiac catheterisation standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.12. Maps of cardiac catheterisation standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Joint procedures

Surgery after hip fracture

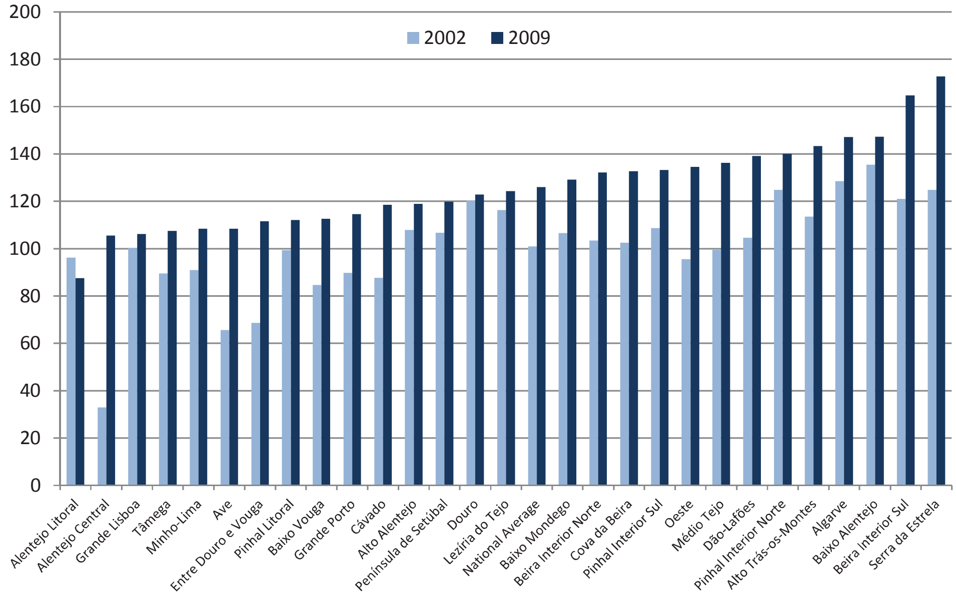
Regarding surgery after hip fracture, a procedure used for "calibration" purposes, the average rate increased by about 25% between 2002 and 2009, while the coefficient of variation was reduced and is, as expected, one of the lowest of all procedures covered under this study (Table 11.8 and Figures 11.13 and 11.14). In 2009, the regions with the lowest values were Alentejo Litoral, Alentejo Central and Grande Lisboa, while the ones with the highest were Baixo Alentejo, Beira Interior Sul and Serra da Estrela. The low variation observed in other countries is thus also present in Portugal.

Table 11.8. Surgery after hip fracture standardised rate per 100 000 population, Portugal, 2002-09

NUTS III	2002	2003	2004	2005	2006	2007	2008	2009
Average of all age groups	101	106	111	109	108	114	120	126
Q10	80	89	91	91	85	90	99	107
Q90	125	123	127	128	125	134	141	147
CV	0.21	0.23	0.18	0.14	0.17	0.17	0.15	0.15

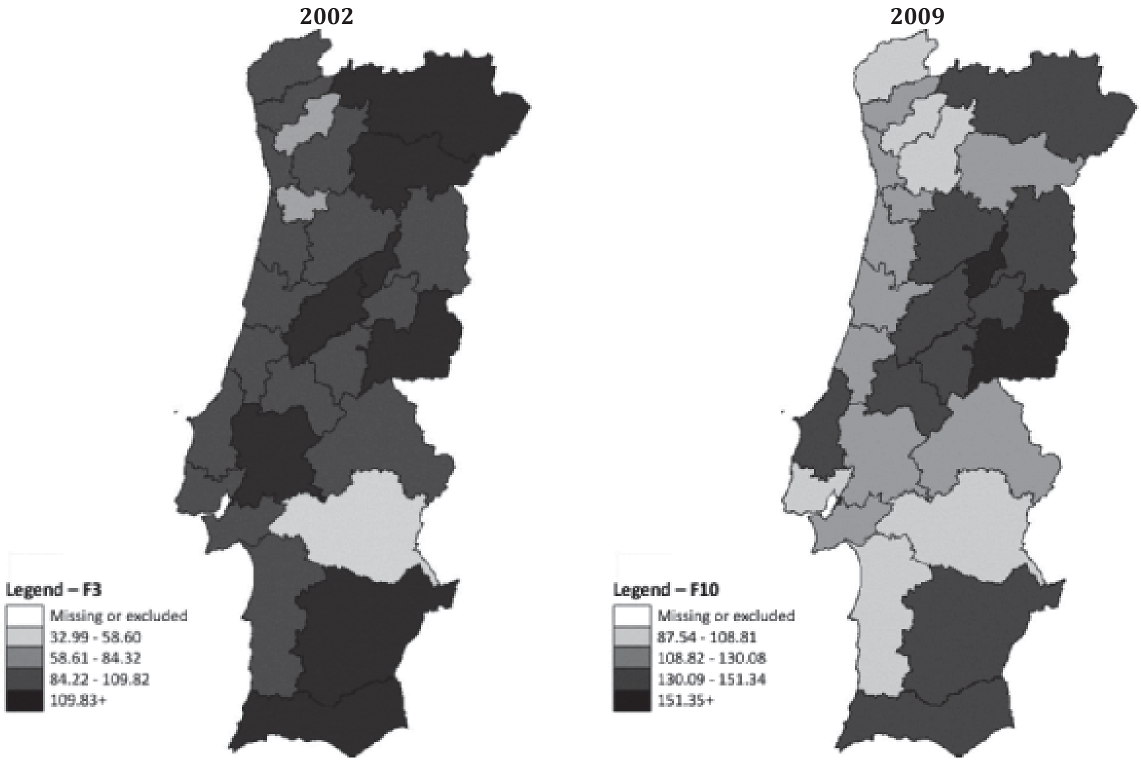
Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.13. Surgery after hip fracture standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.14. Maps of surgery after hip fracture standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Knee replacement

Portugal has a low rate of knee replacement when compared with other OECD countries (OECD, 2013a), but the number of these operations has grown rapidly in a context of population ageing. From 2002 to 2009, the unweighted average rate of knee replacement more than doubled across regions in Portugal (Table 11.9, Figures 11.15 and 11.16). This was accompanied by a reduction in the degree of variation across regions between 2002 and 2006, indicating that the growth rate was particularly rapid in those regions that had low rates. However, since 2007, the geographic variations have widened again, suggesting more rapid growth in those regions that have already high rates.

The regions of Baixo Alentejo and Alentejo Litoral had very low rates of knee replacement in 2002, but following strong and steady growth, they were among the regions with the highest rates in 2009, immediately after the Alto Alentejo region.

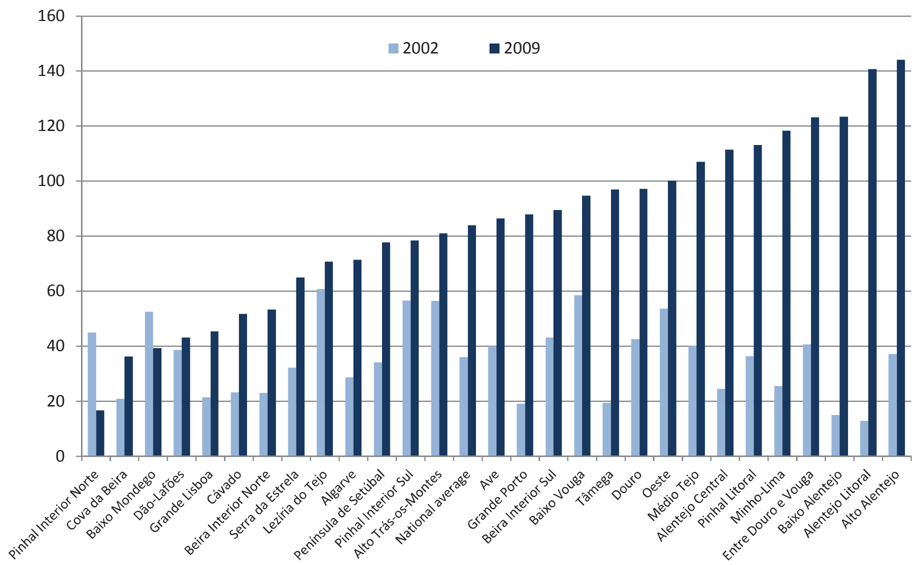
The decrease observed in Pinhal Interior Norte and Baixo Mondego might mean that in both regions most of the people that should receive a knee replacement have already received it or that patients are being moved to the private sector. Further analysis in the changes in the number of orthopedic surgeons in hospitals in the surrounding areas might help to understand the evolution in the waiting list.

Table 11.9. Knee replacement standardised rate per 100 000 population, Portugal, 2002-09

NUTS III	2002	2003	2004	2005	2006	2007	2008	2009
Unweighted average	36	44	55	52	70	70	84	84
Q10	19	22	31	30	47	38	39	42
Q90	57	69	79	76	97	102	121	123
Coefficient of variation	0.4	0.39	0.34	0.34	0.28	0.37	0.39	0.39

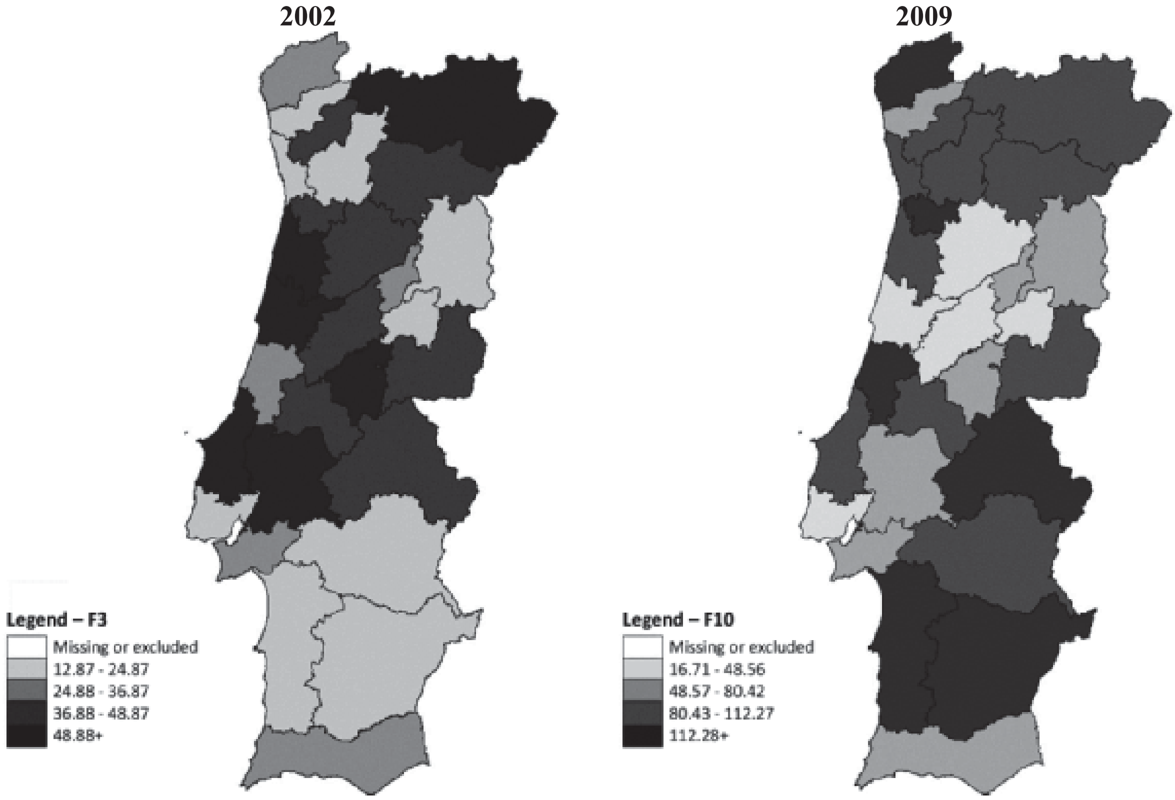
Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.15. Knee replacement standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.16. Maps of knee replacement standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Knee arthroscopy

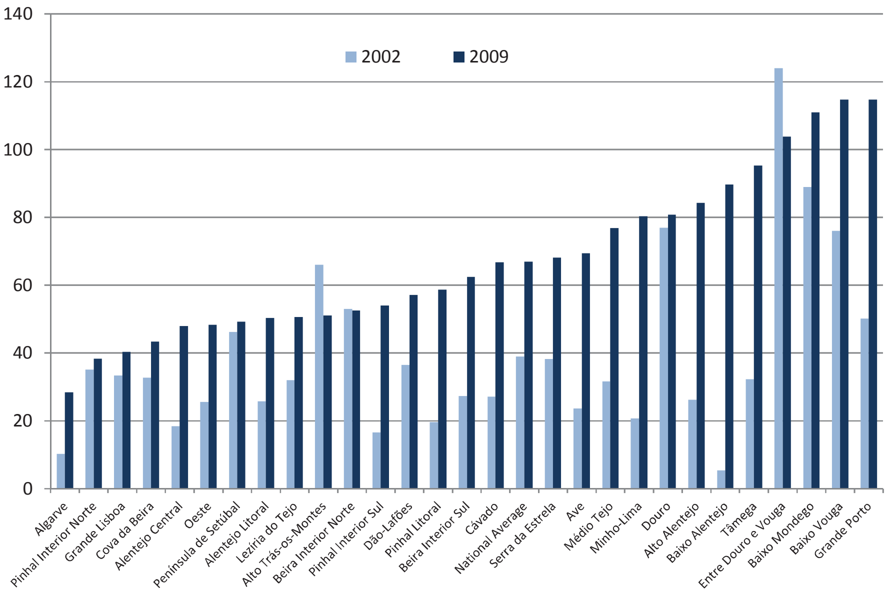
With respect to knee arthroscopy, the unweighted average rate across regions increased by over 70% between 2002 and 2009, while the degree of variations across regions was reduced substantially, indicating a more rapid growth rate in some of the regions that had relatively low rates in 2002 (Table 11.10). This was the case notably in the Algarve region, although the rates of knee arthroscopy remained lower than in other regions in 2009 (Figures 11.17 and 11.18).

Table 11.10. Knee arthroscopy standardised rate per 100 000 population, Portugal, 2002-09

NUTS III	2002	2003	2004	2005	2006	2007	2008	2009
Unweighted average	39	44	52	51	57	65	60	67
Q10	18	23	27	16	29	25	31	42
Q90	76	73	81	84	85	98	97	106
Coefficient of variation	0.67	0.53	0.43	0.51	0.37	0.43	0.45	0.36

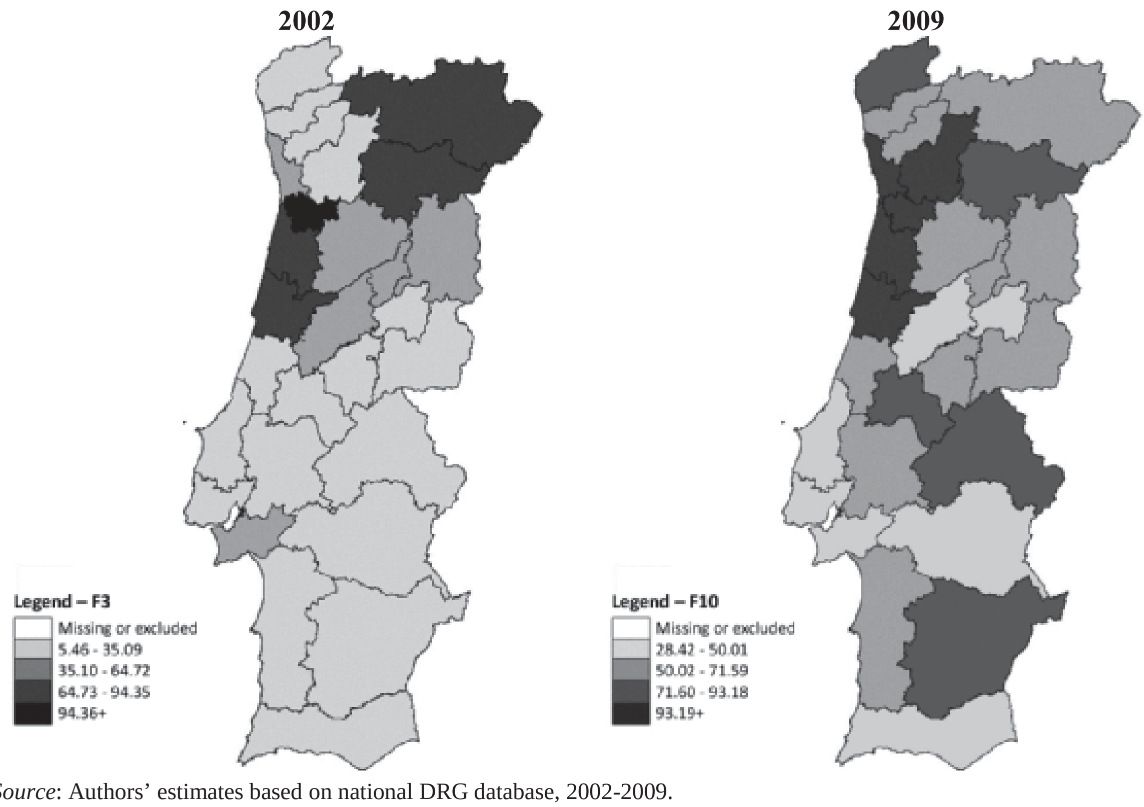
Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.17. Knee arthroscopy standardised rate per 100 000, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.18. Maps of knee arthroscopy standardised rate per 100 000 population, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Gynaecological procedures

Caesarean section

Caesarean sections have been increasing as a proportion of all live births. It is worth mentioning that a noteworthy reform of maternity services took place in 2006. Out of all NHS hospitals, around ten maternity services were closed because they were performing very few deliveries. This might result in less caesarean sections for women living in those regions because they would deliver their babies in hospitals in regions with better trained maternity teams. Nearly 90% of all deliveries in 2009 took place in NHS hospitals, with the remaining 10% occurring in private hospitals. However, there are large variations in caesarean section rates between public and private hospitals: 33% of deliveries in public hospitals were caesarean sections, whereas this proportion reached 66% of deliveries in private hospitals in 2009.

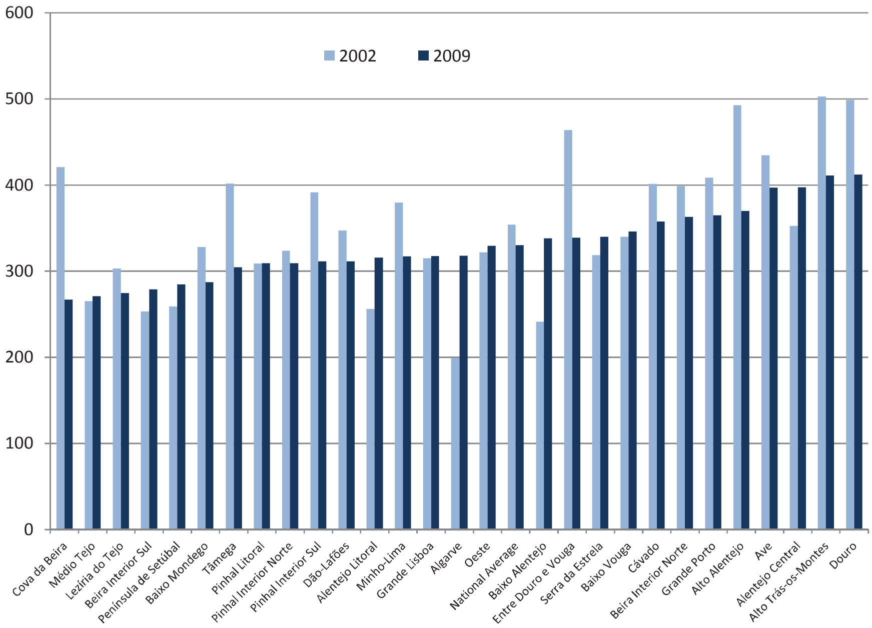
The total number of caesarean sections and total deliveries fell during the study period but the share of caesarean section as a percentage of total deliveries increased. The age-standardised rate of caesarean sections per 1 000 live births, however decreased from 354 in 2002 to 330 in 2009, but peaked at 365 in 2005 (Table 11.11). This implies that the changes in the number of caesarean sections have not been able to compensate for the changes in the number of deliveries over the study period. Douro and Alto Trás os Montes were the regions with the highest rates of caesarean sections in 2009, although the rates in these two regions decreased markedly since 2002 (Figures 11.19 and 11.20). The caesarean section rate also decreased substantially in the Cova Da Beira region, so that it had the lowest rate in 2009. In some regions such as Algarve, caesarean section rates increased greatly between 2002 and 2009, and are now close to the national average.

Table 11.11. Caesarean section age-standardised rate per 1 000 live births, Portugal, 2002-09

NUTS III	2002	2003	2004	2005	2006	2007	2008	2009
Unweighted average	354	355	359	365	358	344	343	330
Q10	255	268	282	280	286	283	285	278
Q90	472	465	471	445	445	416	419	397
Coefficient of variation	0.23	0.23	0.2	0.2	0.19	0.17	0.15	0.13

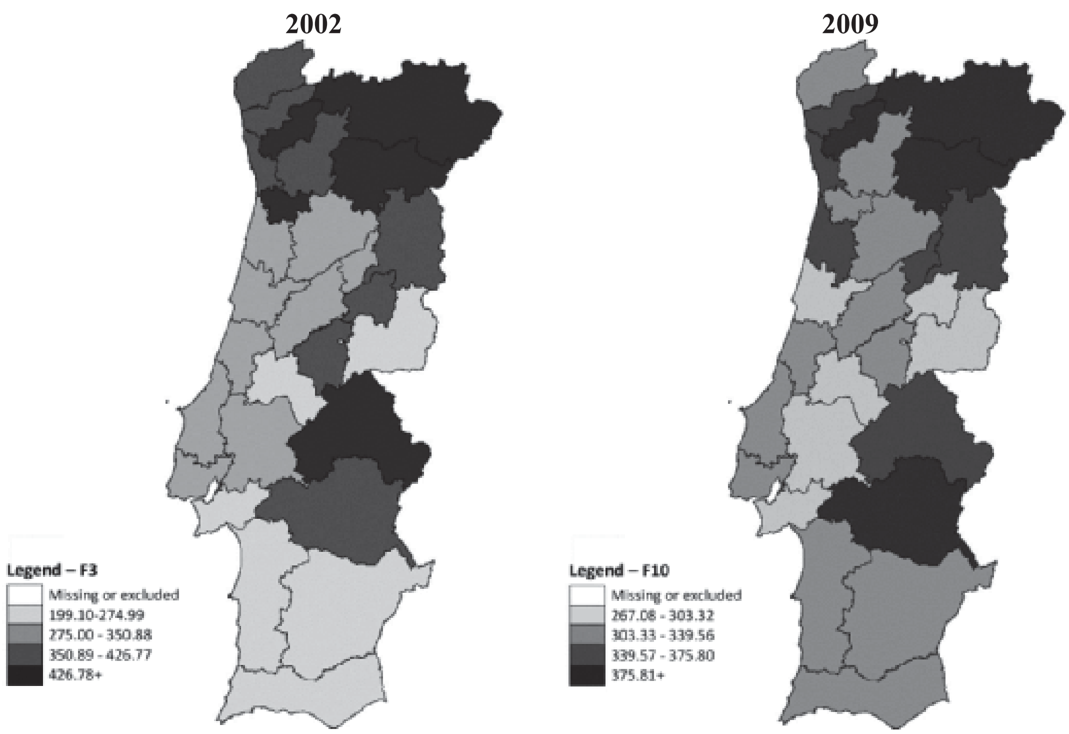
Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.19. Caesarean section age-standardised rate per 1 000 live births, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.20. Maps of caesarean section age-standardised rate per 1 000 live births, by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Since the percentage of total deliveries involving a caesarean section is rising, the reduction in the coefficient of variation during the study period might mean that more women are being offered caesarean sections, which is contributing to the decrease in regional differences. Another contributing factor is the decrease in the number of caesarean sections in the top 10th percentile where there was a decrease in Q90 during the study period.

Since 2007, the rate of caesarean sections is taken into account in assessing hospital performance and for financing purposes, including penalties for hospitals that have rates above what is considered desirable (Valente, 2010).

Hysterectomy

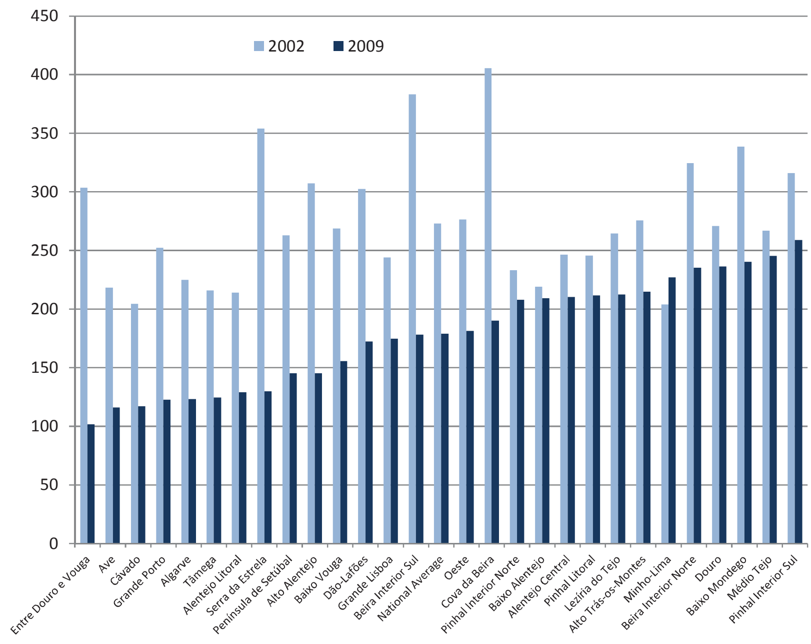
By contrast with the previous interventions, the average rate of hysterectomy has come down significantly in Portugal as in many other OECD countries, particularly since 2007. However, the regional variations have increased, indicating that the reduction has not been uniform across all regions (Table 11.12). There was even a slight increase in the Minho-Lima region between 2002 and 2009 (Figures 11.21 and 11.22). In 2009, the regions with the lowest rates were Entre Douro e Vouga, Ave and Cávado, while the regions with the highest rates were Baixo Mondego, Médio Tejo and Pinhal Interior Sul.

Table 11.12. Hysterectomy age-standardised rate per 100 000 females, Portugal, 2002-09

NUTS III	2002	2003	2004	2005	2006	2007	2008	2009
Unweighted average	273	287	287	275	277	274	255	179
Q10	215	226	217	232	234	207	188	121
Q90	343	342	351	325	327	341	332	238
Coefficient of variation	0.2	0.17	0.17	0.16	0.14	0.19	0.24	0.26

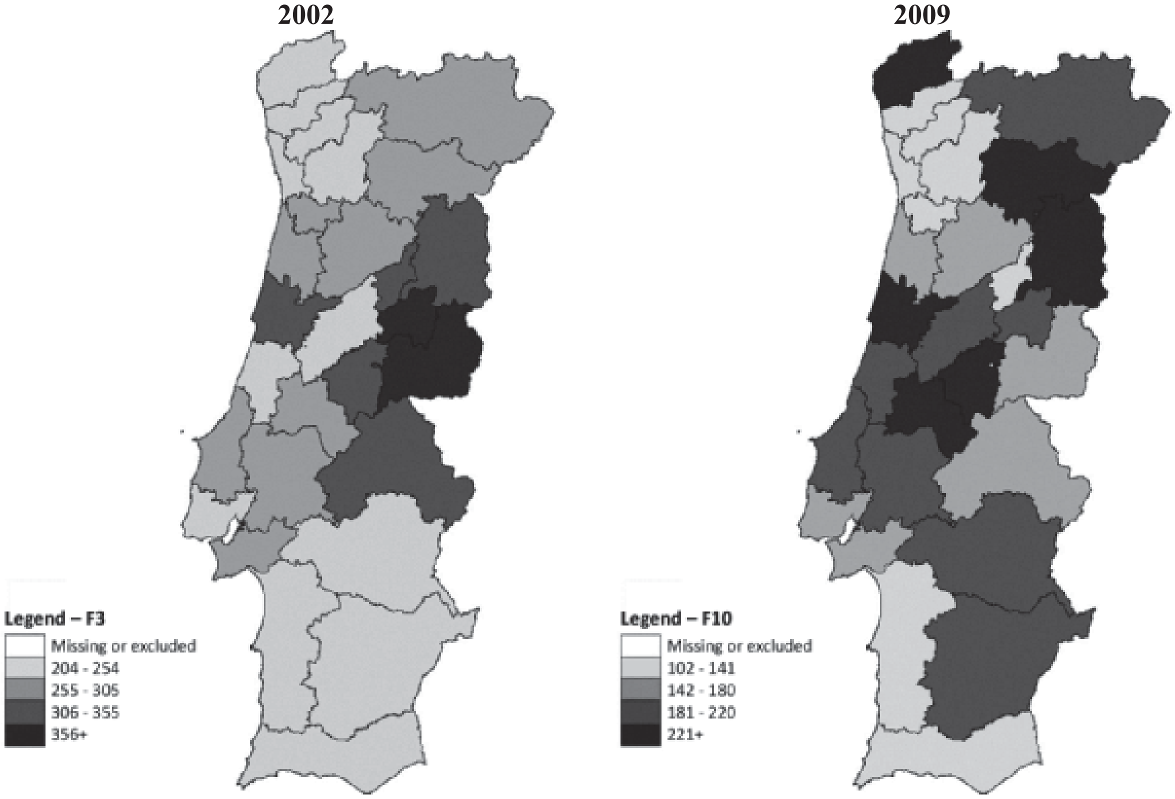
Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.21. Hysterectomy age-standardised rate per 100 000 females by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

Figure 11.22. Maps of hysterectomy age-standardised rate per 100 000 females by geographic region, Portugal, 2002 and 2009



Source: Authors' estimates based on national DRG database, 2002-2009.

11.5. Conclusions

This report has reviewed the evolution of geographic variations in the use of a selected set of health care procedures in Portugal between 2002 and 2009, based on 28 groups of municipalities. The data on utilisation rates have been age- and sex-standardised to remove any effect of different population structures across these different regions and over time. For some of the interventions, there has been a reduction in geographic variations during this period of time, notably for cardiac catheterisation and coronary angioplasty (PTCA) which are used to diagnose and treat ischaemic heart disease, one of the leading causes of mortality in Portugal. The overall increase in PTCA rates combined with the reduction in geographic variations, reflects positive developments in the adoption and access to good clinical practice.

While geographic variations have also decreased for caesarean section rates, the rates as a percentage of deliveries have increased in Portugal but decreased when measured per 1 000 live births. There is evidence that some caesarean sections are not medically required as is the case also in many other OECD countries. Caesarean section rates in Portugal (as in France, Spain and Switzerland) are particularly high in private hospitals, two-times greater than in public hospitals, although caesarean section rates have also been rising in public hospitals where most of the deliveries take place. In response, the Portuguese Ministry of Health has appointed in 2010 a group of experts, with a mandate to: 1) monitor both rates and complications; 2) issue guidelines, particularly regarding the follow-up of low-risk pregnancies; 3) develop guidance for a training programme aimed at health care professionals; 4) propose a plan for both internal and external audits; 5) propose a communication plan for the general population; 6) verify the existence of adequate resources (human and others) in all maternity services; and 7) help implement caesarean section registries and new payment schemes for hospitals.

The number of knee arthroscopies and knee replacements has increased markedly in Portugal between 2002 and 2009, as is the case also in many other OECD countries. While the geographic variations in knee arthroscopies decreased to a certain extent, it still remains very high, and there has been no reduction in the large geographic variations in knee replacement. This means that knee replacement rate has grown as rapidly in regions that had high rates compared with regions that had low rates. With the growing use of this procedure, it is becoming increasingly important to ensure that decisions to perform a knee replacement are based on proper clinical assessment of the potential benefits and risks of the intervention for each patient, and that each patient is properly informed of these potential benefits and risks.

In general, there remain significant geographic variations in the use of different health care procedures in Portugal which cannot be explained by population characteristics. This situation points towards the need to improve access to appropriate care, in order to improve the health outcomes of the Portuguese population.

A National Strategy for Quality in Health Care was launched in 2009 and sets out the goal of improving clinical and organisational quality as well as patient safety (Ordinance 14223/2009). More recently, a national network was created to address the need for continuous improvement in health care quality and to foster a better articulation between the different levels of care.

Minimum volume of activity thresholds should also be taken into consideration in order to reach high-quality standards. This might increase geographic variation if it leads to closing down of small surgical units and their concentration in fewer and bigger hospitals particularly if access problems persist or get worse. People living in the affected region should be offered the procedure but sometimes the hospital will be located further away, which might be a disincentive for people to get the procedure.

There are no decision aids for patients, and patient empowerment is still in its infancy in Portugal. A more systematic collection of information on patient health outcomes, assessed through instruments such as EQ-5D, SF-36, the Oxford Knee Score or the Western Ontario and McMaster Universities Osteoarthritis Index, could also assist in assessing the benefits of different interventions and inform physicians' decision making. In Sweden and England, patient reported outcomes measures are used to study local utilisation rates [see Chapter 14 on United Kingdom (England)].

Even though numerous clinical guidelines have been published by the General Directorate of Health, these have been targeted mainly at the prescription and use of pharmaceuticals. There is a need to develop and monitor the implementation of up-to-date clinical guidelines for diagnostic and surgical procedures, to promote greater harmonisation of medical practices in Portugal. Reasons for non-compliance with the recommended clinical guidelines should be examined closely.

Equity in access to health care is one of the main goals of the Portuguese NHS. If part of the variation observed is the result of barriers to access to care, these barriers need to be identified and measures should be implemented to overcome them.

Knowledge about unwarranted medical practice variations in Portugal is still scarce. More research on this topic, in particular about some of the reasons for these variations, might contribute to improving access to appropriate care for the Portuguese population.

References

Barros, P., S. Machado and J. Simões (2011), "Portugal: Health System Review", *Health Systems in Transition*, Vol. 13, No. 4, pp. 1-156.

Despacho No. 14223/2009 (Ordinance No. 14223/2009), 8 June, *Diário da República*, 2nd series, No. 120, 24 June 2009.

DGS – Directorate General of Health, Direção de Serviços de Informação e Análise. Divisão de Epidemiologia e Estatística. Centros de Saúde e Unidades Hospitalares (2012), "Recursos e produção do SNS: Ano de 2010", Direção-Geral da Saúde, Lisboa, 121 p. available at www.dgs.pt/estatisticas-de-saude/estatisticas-de-saude/publicacoes/centros-de-saude-e-unidades-hospitalares-recursos-e-producao-do-sns-2010.aspx.

INE – National Statistical Institute (2013), "Inquérito aos Hospitais", Lisboa, http://webinq.ine.pt/public/files/inqueritos/InqApresentacao.aspx?Id=1540&inq=IHSP.

INE (2010), Inquérito aos Hospitais 2010, Lisboa.

INE (2009), Estimativas Anuais da População Residente (Annual estimates of the resident population in Portugal), Lisboa.

Mateus, C. (2011), "Portugal: Results of 25 Years of Experience with DRGs", in R. Busse, A. Geissler, W. Quentin and M. Wiley (eds.), *Diagnosis Related Groups in Europe: Moving Towards Transparency, Efficiency and Quality in Hospital*, Open University Press, McGraw Hill, London, pp. 381-400.

OECD (2013a), *Health at a Glance 2013: OECD Indicators*, OECD Publishing. http://dx.doi.org/10.1787/health_glance-2013-en.

OECD (2013b), *OECD Health Statistics 2013*, Paris: OECD Publishing. http://dx.doi.org/10.1787/health-data-en.

OECD (2012a), *Health at a Glance: Europe 2012*, OECD Publishing. http://dx.doi.org/10.1787/9789264183896-en

OECD (2012b), OECD Health Systems Characteristics Survey, 2012, www.oecd.org/els/health-systems/characteristics.htm.

Valente, M.C. (2010), "Contratualização em contexto hospitalar" [The Portuguese hospitals purchaser-provider model], *Revista Portuguesa de Saúde Pública*, Vol. Temático No. 9, pp. 25-39.

Potential of geographical variation analysis for realigning providers to value-based case. ECHO case study on lower-value indications of C-section in five European countries (Work 2)

Sandra García-Armesto
Ester Angulo-Pueyo
Natalia Martínez-Lizaga
Céu Mateus
Inês Joaquim
Enrique Bernal-Delgado

European Journal of Public Health:
Feb; 25 Suppl 1: 44-51

2015

DOI: 10.1093/eurpub/cku224.

European Journal of Public Health, Vol. 25, Supplement 1, 2015, 44–51

doi:10.1093/eurpub/cku224

Potential of geographical variation analysis for realigning providers to value-based care. ECHO case study on lower-value indications of C-section in five European countries

Sandra García-Armesto[1,2,3], Ester Angulo-Pueyo[1,3], Natalia Martínez-Lizaga[1,3], Céu Mateus[4], Inês Joaquim[4], Enrique Bernal-Delgado[1,3]* on behalf of the ECHO Consortium

1 Health Services and Policy Research Unit, Health Sciences Institute in Aragon (IACS) IIS Aragon, Zaragoza, Spain
2 Agency Aragon R+D Foundation (ARAID), Zaragoza, Spain
3 Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Spain
4 National Public Health School (ENSP), Nova Universidade de Lisboa, Portugal

Correspondence: Sandra García-Armesto, Institute for Health Sciences in Aragon, IIS Aragon, Zaragoza, Spain, Tel: +34 976713992, e-mail: sgarciaarm.iacs@aragon.es

*The members of the ECHO Consortium are listed in the acknowledgments.

Background: Although C-section is a highly effective procedure, literature abounds with evidence of overuse and particularly misuse, in lower-value indications such as low-risk deliveries. This study aims to quantify utilization of C-section in low-risk cases, mapping out areas showing excess-usage in each country and to estimate excess-expenditure as a proxy of the opportunity cost borne by healthcare systems. **Methods:** Observational, ecologic study on deliveries in 913 sub-national administrative areas of five European countries (Denmark, England, Portugal, Slovenia and Spain) from 2002 to 2009. The study includes a cross-section analysis with 2009 data and a time-trend analysis for the whole period. Main endpoints: age-standardized utilization rates of C-section in low-risk pregnancies and deliveries per 100 deliveries. Secondary endpoints: Estimated excess-cases per geographical unit of analysis in two scenarios of minimized utilization. **Results:** C-section is widely used in all examined countries (ranging from 19% of Slovenian deliveries to 33% of deliveries in Portugal). With the exception of Portugal, there are no systematic variations in intensity of use across areas in the same country. Cross-country comparison of lower-value C-section leaves Denmark with 10% and Portugal with 2%, the highest and lowest. Such behaviour was stable over the period of analysis. Within each country, the scattered geographical patterns of use intensity speak for local drivers playing a major role within the national trend. **Conclusion:** The analysis conducted suggests plenty of room for enhancing value in obstetric care and equity in women's access to such within the countries studied. The analysis of geographical variations in lower-value care can constitute a powerful screening tool.

Introduction

Background

In serving an aging population within an ever-evolving technological environment, health systems' affordability, quality and equity should be encompassed to achieve sustainability. In addition, European Governments are currently faced with the need to adjust expenditure to acute budget contractions in line with the public deficit threshold enforced in the Eurozone.[1] There is a clear need for generating tools and shaping solid analytical methods that could support decision making in optimizing the use of available resources.[2]

The European Collaboration for Healthcare Optimization (ECHO) project aimed at exploring such a possibility on the basis of the analysis of geographical medical practice variations. The conceptual approach was very simple: Health systems bear substantial opportunity cost in using interventions deemed lower-value. Quantifying the utilization of this type of care and its systematic variation across policy-relevant geographical units could offer 'at glance' insights into local room for realignment into value-based provision of care. In addition, geographical differences in residents' exposure to lower-value care might be signalling inequalities in access to quality and safe care that should be tackled.

Several countries (Australia[3,4], Canada[5] and New Zealand[6]) have already taken steps to minimize the deployment of resources on lower-value care. Commissioning bodies around England, such as the Croydon Primary Care Trust, pioneered the production of guiding lists of lower-value procedures[7] and the Rightcare NHS initiative has continued the task of supporting the Clinical Commissioning Groups in limiting the commissioning of such services.[8]

Based on those experiences reported in the literature, we identified a small set of elective interventions considered lower-value care, with a view to describing and mapping out unwarranted variations in their utilization across geographical areas in five countries. The set included: (i) procedures superseded by more cost-effective alternatives (non-conservative breast cancer surgery, hysterectomy in non-oncological conditions); (ii) interventions where there are defined types of patients for whom evidence of value is doubtful (prostatectomy in benign prostatic hyperplasia, C-section in low-risk pregnancies and deliveries) and (iii) relatively ineffective procedures prone to overuse (adenoidectomy and tonsillectomy). The results of this exercise have been extensively reported in the corresponding five Country Atlases on lower-value care available from the ECHO project webpage.[9]

This article will focus on C-section in low-risk pregnancies and deliveries as a case-study to unfold the ECHO approach and discuss the potential usefulness of combining variations and value-for-money metrics, using national benchmarks within an international perspective. C-section is considered a highly effective procedure in avoiding maternal and child mortality at birth as well as complications derived from foetal distress. However, in the last decade, literature abounds with evidence of overuse, and particularly misuse, in lower-value indications such as low-risk pregnancies and deliveries.[10–13] According to these analyses, areas with higher usage rates 'perform the intervention in medically less appropriate populations—that is, relatively healthier births—and do not see improvements in maternal or neonatal mortality'. Several non-medical factors have been consistently associated with higher caesarean rates.[13,14] They include: provider density, the capacity of the local health care system and malpractice liability.

The World Health Report 2010 'Health systems financing: the path to universal coverage' addressed the issue, supported by two background papers examining the determinants of C-section use in developed countries and estimating the global economic impact of unnecessary interventions.[15,16] They even propose an empirical threshold in the intensity of use, suggesting that 'Countries with C-section rates below 10% were considered to show under use, while countries with rates above 15% were considered to show overuse'.

Along the same lines, OECD's cross-country comparisons over the years confirm a general increasing trend—the average caesarean rate among OECD members grew from 14% of all births in 1990 to nearly 20% in 2000 and 26% in 2009—notwithstanding remarkable variations across countries—in 2009 the range across OECD countries went from 14% in the Netherlands to almost 40% in Italy.[17] The ECHO project aimed at specifically targeting the utilization of lower-value indications of C-section, that is, interventions performed in women without documented risk factors for complications in pregnancy or during delivery which may advise C-section over vaginal delivery. Total C-section rates per circumscription were also examined to calibrate results.

This study aims to quantify the utilization of C-section in low-risk deliveries and pregnancies and identify areas showing excess-usage in each country's statutory system to map out where decision makers' priorities may lie in realigning obstetric care into value-based provision. Excess-expenditure derived from such utilization rates is estimated to offer a rough proxy of the opportunity cost borne by systems.

Methods

Design

Observational, ecologic study on the deliveries occurring in five European countries (Denmark, England, Portugal, Slovenia and Spain) from 2002 to 2009. The study includes a cross-section analysis with 2009 data and a time-trend analysis for the whole period. Data for Slovenia were only reliable for 2005–9 due to changes in coding practices.

Population

All deliveries within the statutory health system in Denmark, England, Portugal, Slovenia and Spain, assigned to a woman's area of residence.

Unit of analysis

Sub-national administrative areas ($n = 913$) relevant in health policy and care planning within Denmark [98 areas (Municipalities or Kommuners); mean pop. 56 000], England [326 areas (Local Authorities); mean pop. 159 000] Portugal [278 areas (Municipalities or Concelhos); mean pop. 36 000], Slovenia [12 areas (statistical regions); mean pop. 169 000] and Spain [199 areas (Healthcare Areas); mean pop. 234 000].

Main endpoints

Age-standardized utilization rate of C-section in low-risk pregnancies and deliveries (sRC-sLR) per 100 deliveries for each relevant boundary in each country. The indicator was defined by using C-section procedure codes and excluding all those cases that included diagnosis codes compatible with conditions or delivery situations increasing the risk of complications for a vaginal delivery (for instance, mother's health, multiple gestation, malposition and malpresentation of foetus, placenta previa and haemorrhage, disproportion and foetal distress.[18] As a summary measure of variation, the classical statistics ratio of variation between extremes or Extremal Quotient (EQ95-5) was calculated together with the Systematic Component of Variation (SCV).

Secondary endpoints

The estimated excess-cases per geographical unit of analysis in two scenarios of minimized utilization. Benchmarks were set using the minimal standardized utilization rates observed within each country. To allow some sensitivity, two scenarios were designed: the first took as a benchmark the behaviour of areas in the first quartile of the national distribution of utilization (p25); the second, more demanding, set the desirability standard at the 10th percentile (p10).

Excess-cost per country was estimated as the difference between unit costs of uncomplicated C-sections and that of uncomplicated vaginal delivery (country tariffs) multiplied by the estimated excess-interventions. Only in the case of Denmark and Slovenia, where national tariffs for the period of analysis were not publicly available, did the estimation of the marginal unit costs rely on the detailed calculations performed by Gibbons et al.[16] for the WHO 2010 report. The resulting figure is intended as a raw proxy of marginal expenditure/resource deployment on lower-value indications of C-section.

Statistical analysis

Primary analyses focused on eliciting the percentage of delivery by C-section in low-risk cases over total deliveries in each area and its systematic variation across the corresponding country. For that purpose, the component of systematic variation (CSV) and its confidence intervals were estimated.

'Age-standardized rates of C-section in low-risk cases per 100 deliveries' (sRC-sLR) were calculated for each area using the direct method, taking as a reference the age distribution of the mothers in the five countries.

For lower-value care, excellence is about minimizing utilization. Therefore, the expectation built in calculating excess-cases was the 'minimal' rate, that is, the lowest rate in the country; given the risk of small numbers and the statistical impact of over-dispersion, it was thought reasonable to set a 'generous' minimum, such as the rate corresponding to the lowest quartile of the distribution across areas. Nevertheless a second, more demanding, scenario was also considered, involving the bottom 10% of the distribution (p10).

The adoption of national, rather than cross-country, benchmarks was intended to enhance the acceptability of goals within local organizations. As the ultimate objective was to provide decision makers with tools to realign providers to value-based provision, a national benchmark could help in setting more attainable minimization targets, keeping the international comparison context to raise questions about the factors underpinning cross-country differences and eventually setting aspirational international benchmarks.

Nine age-specific minimal utilization rates (5-year groupings for women between 15 and 55 years old) were obtained by aggregating deliveries and interventions across those areas in the lowest quartile (or percentile 10, in the alternative scenario). 'Expected cases' were obtained by indirect standardization: the number of interventions that should be expected among women giving birth in an area, were its rate the minimal utilization rate.

'Excess-cases' were calculated as the difference between the actual number of cases within each boundary and those expected—obtained from the minimization to the benchmark utilization rate (p25 or p10, depending on the scenario).

Two measures of variation were calculated: (i) $EQ_{95\text{-}5}$ as the ratio of utilization rates in areas in the 95th and 5th percentiles of the distribution, (ii) SCV, focused on eliciting the proportion of the observed variation exceeding that expected by chance, which was estimated following a two-step hierarchical model[19]—for details on these estimation methods see the *ECHO Handbook*[20] and the paper on Potentially Avoidable Hospitalizations in this special issue. Finally, 'trends in average standardized utilization rates and SCV' in each country were analysed from 2002 to 2009, taking as a reference the age distribution of mothers in 2002.

'Area-specific total C-section age-standardized rate per 100 deliveries (sRC-s) and its variation' across countries was also estimated following the methods described earlier. The intensity of C-section usage was considered as a potential factor contributing to explain the use of C-section in the lower-value indications. A pairwise correlation between the two variables for all 913 areas in the five countries and per country was tested using the Pearson product-moment correlation coefficient.

Data sources

The ECHO data warehouse containing all hospitalizations within the statutory systems in the five countries for the period of analysis allocated to the relevant geographical units. For marginal unit cost of procedures per country the sources were: 2008/09 Admitted Patient Care Mandatory Tariff (England), Ministerio da Saude Portaria no. 132/2009 de 30 de Janeiro (Portugal), Norma Ministerio APv25.0 datos 2008 (Spain) and Denmark and Slovenia: WHO 2010 report background paper 30 (Slovenia).[16]

Results

Table 1 summarizes utilization rates and variation statistics per country for 2009, the last year in the period analysed, both for total C-section interventions (sRC-s) and the lower-value indication (sRC-sLR). Average crude and standardized rates for C-section varied widely across the five countries; Portugal yielded the highest rate (32% of all births), well ahead of its immediate follower, England with 26%. The lowest rate, 19%, was found in Slovenia. However, this ranking would not hold for sRC-sLR, with Portugal at the bottom, showing 2% of deliveries by the lower-value indication and Denmark at the top, close to 10%—doubling Slovenian and English rates and three folding the Spanish.

Table 1 Utilization of C-section intervention (total and lower-value indication) and its variation in the ECHO countries. Year 2009

	Denmark	England	Portugal	Slovenia	Spain
Deliveries	61 856	636 972	84 638	20 021	403 545
C-section					
Number of interventions	13 489	156 520	26 902	3638	83 426
cRate per 100 deliveries	21.81	24.57	31.78	18.17	20.67
sRate per 100 deliveries	22.16	25.49	32.79	18.97	21.43
$EQ_{95\text{-}5}$	1.30	1.40	1.54	1.50	2.44
SCV	0.01	0.02	0.22	0.05	0.06
C-section LRD					
Number of interventions	5356	26 982	1140	938	9287
% of total C-section	39.71	17.24	4.24	25.78	11.13
cRate per 100 deliveries	8.66	4.24	1.35	4.69	2.30
sRate per 100 deliveries	9.56	4.30	1.56	5.44	2.74
sR_min	3.80	0.00	0.00	2.86	0.00
sR_max	19.00	13.56	15.58	10.11	14.03
sR_p5	5.35	1.90	0.00	2.86	0.14
sR_p10	6.14	2.21	0.00	2.91	0.32
sR_p25	6.81	2.69	0.00	3.10	0.88
sR_p50	8.81	3.79	0.20	4.16	1.97
sR_p75	11.02	4.98	1.34	6.71	3.61
sR_p95	16.88	8.60	7.95	10.11	8.32
$EQ_{95\text{-}5}$	3.16	4.51	–	3.53	41.03
SCV	0.13	0.24	3.07	0.29	1.45

C-section LRD = interventions in low-risk deliveries and pregnancies; cRate = Crude rate; sRate = age standardized rate taking as reference the age distribution of mothers in the five countries; SR_px = the sR corresponding to percentile x of the distribution of standardized utilization rates across areas in the country; EQ_{95_5} = extremal quotient, ratio between sR_p95 and sR_p5; SCV = systematic component of variation

The proportion of total C-sections that could be deemed lower-value care in each country seemed uncorrelated with the size of sRC-s' size: Denmark showed the highest proportion, 40% of the C-sections performed corresponding to lower-value indications, while in Portugal (the largest sRC-s) only 4% of the interventions were performed in low-risk cases; in Slovenia (the lowest sRC-s) almost 26% of the C-sections were lower-value. No statistical correlation between area-specific sRC-s and sRC-sLR was found across the 913 areas analysed. Country-specific analyses (excluding Slovenia, given the low number of units of analysis) found only a weak correlation (Pearson coefficients around 0.3 with $P < 0.01$).

Regardless of their average rate, variation in lower-value utilization of C-section across the territory seems to be noteworthy in all countries. Examining $EQ_{95\text{-}5}$, Spanish healthcare areas' sRC-sLR ranged between almost null for those below the 5th percentile to close to Danish levels for those above the 95th percentile—more than 40-times larger. In England, women living in those areas with the highest rates were 4.5 times more likely to deliver by the lower-value procedure than residents in Local Health areas with the lowest rates. Likewise, women could bear a 3.5 times higher probability of C-sLR depending on where they lived in Slovenia and three times higher in Denmark. In Portugal, the area-specific sRC-sLR remained zero even at percentiles 10th and 25th of the distribution, thus $EQ_{95\text{-}5}$ could not be calculated.

Nevertheless, the systematic component of this observed variation across the 913 areas shows moderate values in Denmark (13% above randomly expected variation) and relatively high in England and Slovenia (24 and 29% respectively), whereas in Spain and Portugal, it reached several times the variation expected by chance.

Regarding total C-section, the probability of this type of delivery increased between 30% in Denmark and 2.5 times in Spain depending on the women's area of residence, but the systematic component of such variation was almost negligible in all countries, denoting relatively homogeneous behaviour across areas. The exception was Portugal, where 22% of the observed variation could not be deemed random.

Figure 1 compares area distributions according to their sRC-sLR and sRC-s across countries. On the left-hand side, areas are represented by their actual rates; on the right, these have been normalized (log transformed and centred on the mean) to overcome the distorting scale effect due to country differences in rate size; the length of the clouds shows how the behaviour in the areas regarding C-section is relatively similar in all countries, with the dots concentrated around the average; only Spanish areas are slightly more stretched along the scale. However, the clouds representing sRC-sLR are clearly elongated in all countries but Denmark, where they spread evenly across a wider range of values (that is, intensities of use of the lower-value indication). The extreme cases are Portugal and Spain where the plot almost takes a linear shape when normalized.

The evolution of rates and variation in each country over time is represented in figure 2. Total C-section rates have increased slightly over the period, with Portugal showing a relatively larger increase, already departing from a higher level of utilization. The sRC-sLR was also relatively stable over the period, showing a slight downward trend in England and Spain (where systematic variation has also decreased, particularly in England) and upwards in Denmark (together with the systematic variation across the country).

The benchmarking against the minimal sRC-sLR was conducted to estimate excess-cases per area, using 2009 data. Figure 3 maps out the results for the conservative scenario (minimal rate = rate in those areas within the lowest quartile of utilization in each country, p25).

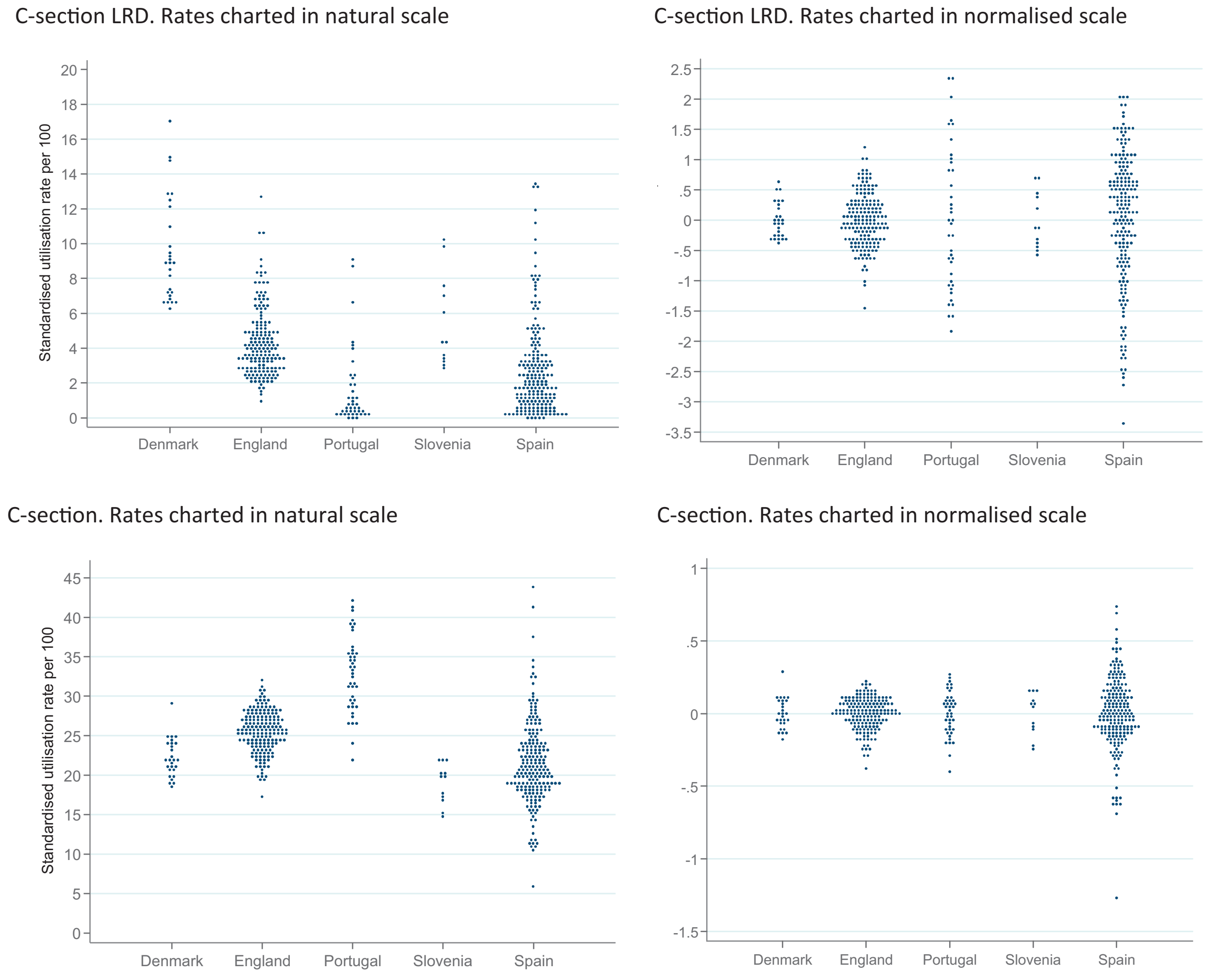


Figure 1 Cross-country comparison of standardized rates of total and lower-value indications of C-section per 100 deliveries per geographical area, year 2009. Each dot represents the relevant administrative area in the country (Municipality in Denmark and Portugal, Local Authority/Healthcare Area for England and Spain and Region in Slovenia). The y-axis charts the area-specific sR per 100 deliveries. The figure is built on the total number of deliveries and C-sections within the statutory health system in 2009 in those countries. Graphics on the right-hand side of the figure represent utilization sRs normalized for ease comparison of degree of variation across countries (log transformed centred on the mean)

Areas are coloured in darker shades the larger the amount of excess-cases they show. For ease of map reading, they have been clustered into quartiles of excess-cases; quartile cut-off points differ from country to country, following the local distribution of cases (affected both by differences in the basal number of deliveries—population—and the magnitude of the minimal rate in each country).

In Denmark, areas above p25th of utilization go from less than 1 to 123 excess-cases, the wider range corresponding to the higher quartile of overuse. English Local Authorities range from 1 to 322 excess-cases of the lower-value C-section, likewise the Spanish healthcare areas; in Slovenian regions excess-cases could reach up to 174, and the larger oscillations are again found in the highest quartile. In Portugal, the range within the 3 first quartiles of overuse is remarkably narrow; the fourth, on the other hand, jumps from 6 to 139.

The maps identify those areas with a higher potential to decrease lower-value obstetric care in each country. Geographical patterns are relatively scattered, with the exception of Denmark where overuse seems to be particularly high in the eastern municipalities of the country.

Table 2 shows the overall country estimation of excess-cases for the two scenarios (p25 and p10), as well as the estimated excess-cost derived. In Spain and Portugal more than 80% of the C-sections in low-risk cases could be avoided if all the areas realigned to the lowest intensity of use in the country. In Slovenia, the potential decline amounted to 38–76% of interventions, depending on the scenario, while English standards set 48–57% of the cases as excess-use. Danish higher basal utilization rates with smaller variation across municipalities resulted in the national benchmark quantifying excess-cases in <30% of the interventions performed in low-risk cases (37% in the most demanding minimization scenario).

Discussion

The aim of this article was to show the type of information that the metrics of geographical variation applied to lower-value care could provide and discuss how useful it could be for decision makers in identifying opportunity cost and local room for enhancing value-based care. The findings show that C-section is widely used in all examined countries at levels that would be considered overuse by the WHO's proposed threshold of 15% of deliveries[16], with Portugal heading the ranking and Slovenia closing it. Furthermore, the analysis provided a measure of how this utilization was actually spread within each country and how systematic the observed variation was. The results suggest that, with the exception of

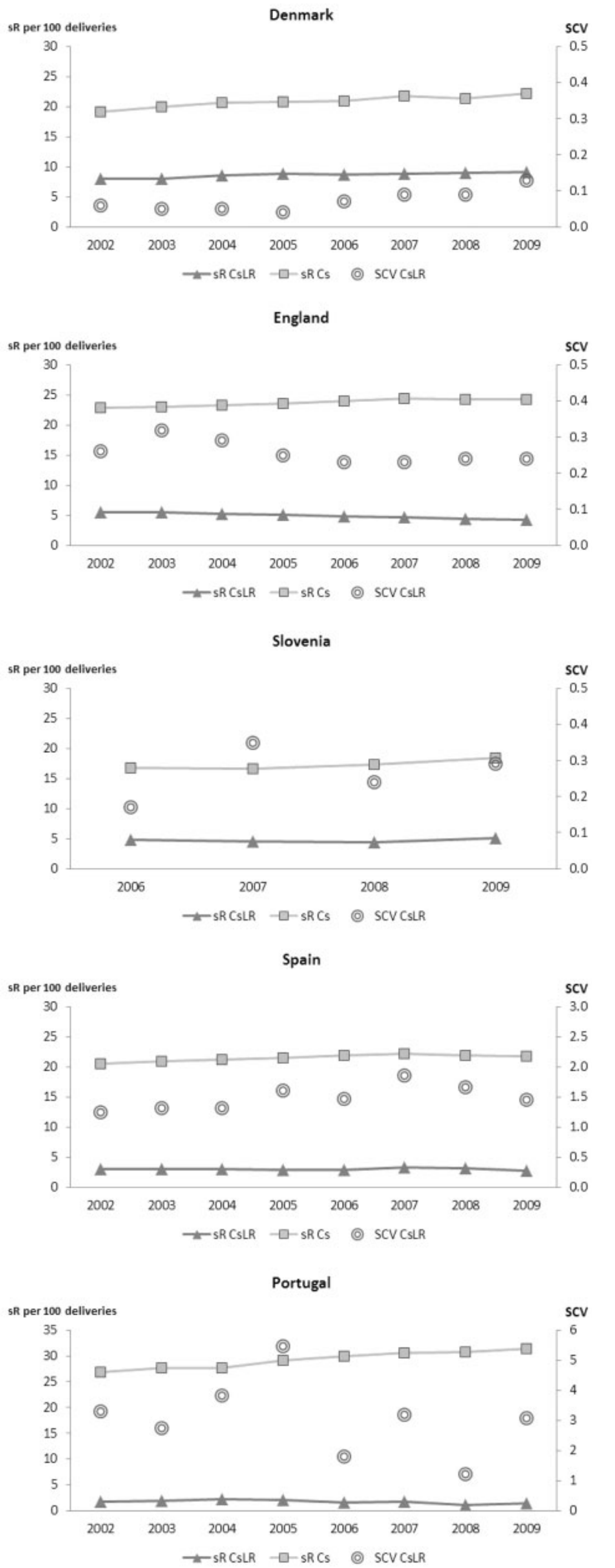


Figure 2 2002–9 trends in average age-standardized percentage rates of deliveries by lower-value C-section (sR CsLR) with their degree of systematic variation (SCV CsLR) and total C-section rates (sR Cs) in ECHO countries. The SCV is plotted on the secondary y-axis on the right; it corresponds to variation of area-specific. sR CsLR across the country for each year

Portugal, there are no systematic variations in intensity of use of the procedure across areas in the same country, despite the differences observed.

However, when focusing on C-section performed in low-risk pregnancies and deliveries, the ratios of utilization within a country, and the proportion of this variation exceeding that could be randomly expected, escalated everywhere. The behaviour described has proven fairly stable over the period of analysis (2002–9), pointing to consolidated practices in each country. In addition, the intensity of utilization of lower-value C-section draws a very different picture for the international comparison than that offered by overall C-section (with Denmark showing the highest rates and Portugal the lowest). This stresses the importance of being specific in defining the type of care that has been deemed lower-value and, thus, is a candidate for measures to minimize a population's exposure and the resources devoted to it. The marked differences in both average and minimal utilization rates (national benchmarks) of lower-value interventions detected across countries, can flag up for decision makers the relative room for improvement in their context.

Within each country, the scattered geographical patterns of lower-value C-section intensity of use speak for local drivers playing a major role within the national trend. The maps clearly identify those areas above the national minimal utilization rate, where further inquiry is warranted to seek the factors underpinning such findings. Based on routinely collected information, they can constitute a powerful screening tool for the system. In addition, the quantification of excess-interventions provides decision makers with a better grasp of the dimension of the phenomenon and its potential severity for the exposed population.

According to the literature[13–15,21], drivers of such intensity of utilization might ostensibly lie on the supply side, relating to provider density, the capacity of the local health care system, malpractice pressure, management of induced labour and quality of obstetric care. All of them are susceptible to intervention by decision makers in order to realign providers' incentives to deliver value-based care. The typical measures will include implementation of clinical guidelines/protocols to improve quality of care standards, revision of providers payment systems, guidelines for commissioners and reorienting capacity.

For the approach to be useful, the eventual intervention to minimize lower-value care should be guided by an understanding of the local situation and the specific factors leading to it in each context.[2] The necessary discussions among relevant stakeholders at local level could be easily supported with similar analyses of utilization rates and excess cases per hospital and even department. Those areas showing the lowest utilization rates should also be explored to understand the keys of those findings and eventually learn lessons that could be shared.

Limitations of the analysis

Information bias is always a risk in dealing with data from hospital administrative databases. Being routinely reported for a variety of purposes core to the system, their completeness and quality have been perfected over the years. However, coding practices might still have an impact; for instance, certain diagnosis codes favouring C-section—such as disproportion or foetal distress—might be used with differential criteria. The under-declaration of certain maternal conditions advising a C-section might also be an issue that would play in the opposite direction.

The choice of national rather than international benchmarks in building the excess-utilization scenarios partly overcomes the risks that such an information bias may pose to their relevance for local decision makers. The issue of whether differences stem from hospital malpractice in coding the diagnoses used to identify low-risk deliveries still remains. Nevertheless, since the analysis is intended as a screening tool to guide decision makers in engaging in

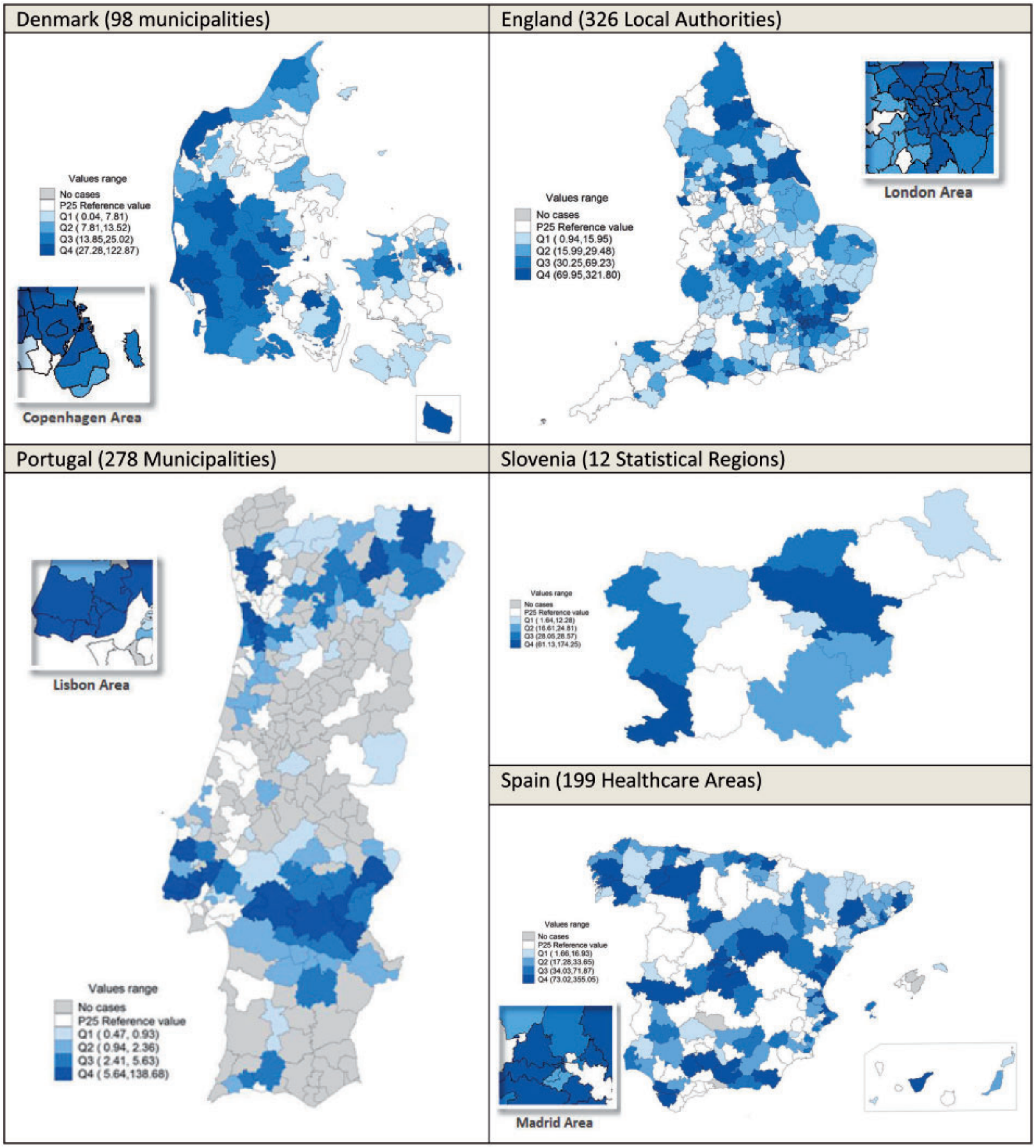


Figure 3 Excess-cases of C-section in low-risk pregnancies and deliveries, by country. Year 2009, minimization to the lowest rate of utilization in the country (benchmark: areas below the 25th percentile). Maps represent the excess-interventions of lower-value caesarean section in each area. The darker the shade the larger is the difference between the observed number of interventions and the benchmark (number of interventions if the area behaved like those with the lowest utilization in the country—below the 25th percentile). Areas are clustered into quartiles according to their level of excess-interventions (Q1 to Q4). Legend provides each quartile's range of values within the country

discussion with providers, this type of explanation is likely to be fully unveiled, unfolding a potential for enhanced reporting.

Another limitation lies in the fact that the ECHO database does not contain information about deliveries outside the statutory system in each country. In this respect, the analysis has adopted the public decision maker perspective, targeting activity under the responsibility of those providers, who are accountable to such decision makers. If a significant proportion of obstetric care were delivered outside those boundaries, women's exposure to lower-value care would be invalid. If such a proportion were very different across countries, the observed differences in the intensity of utilization would be biased. However, comparing the number of deliveries in the database (Table 1) with those reported in the corresponding national statistics, >95% of the activity took place within the scope of the ECHO in Denmark and England, around 90% in Portugal and Slovenia and some 80% in Spain. Thus, the findings are likely to provide a fair picture of women's exposure to C-section and its lower-value indication in those countries.

Finally, it should be underlined that the findings presented are based on ecologic analyses—individual data aggregated at a certain geographical level, which becomes the unit of analysis. Thus, the correct interpretation highlights the risk of being exposed to lower-

Table 2 Estimated opportunity cost by country on the basis of the marginal unitary cost of excess-lower value C-sections in 2009

	Number of Cs LRD	Marginal unitary cost[a]	Minimizing scenario p25		Minimizing scenario p10	
			Excess-CsLR (%)	Excess-cost	Excess-CsLR (%)	Excess-cost
Denmark	5356	DK 8604.48	1456 (27.2)	DK 12 528 123	1965 (36.7)	DK 16 907 803
England	26 982	£1202.00	12 895 (47.8)	£15 499 790	15 492 (57.4)	£18 621 384
Portugal	1140	€ 449.78	951 (83.4)	€ 427 741	1022 (89.7)	€ 459 675
Slovenia	938	€ 1421.57	357 (38.1)	€ 507 500	714 (76.1)	€ 1 015 001
Spain	9287	€ 1318.34	7597 (81.8)	€ 10 015 429	8448 (91.0)	€ 11 137 336

Two scenarios depending on the minimal utilization benchmark: the rate among areas below the 25th percentile of utilization in the country or those below the 10th percentile.
a: Difference between tariffs for C-section without complications and vaginal delivery without complications, local currency.

value care for the women living in a certain area (as opposed to the risk for an individual patient).

Conclusions

The analysis conducted suggests that there is plenty of room for enhancing value in obstetric care while assuring equity in women's access to such care within ECHO countries. Further analysis on institutional factors underpinning overuse of lower-value C-section at small area level, as well as local social, organizational and budgetary contexts, should serve as a basis for recommendations to guide relevant decision makers in tackling this allocative inefficiency. The analysis of geographical variations in lower-value care can constitute a powerful screening tool for the systems. This sort of analysis is promising in offering the kind of information that could prompt clinicians and decision makers to deliver change.

Acknowledgements

ECHO Consortium: IACS's team (Bernal-Delgado E, García-Armesto S, Martínez-Lizaga N, Seral M, Estupiñán F, Comendeiro M, Angulo-Pueyo E, Ridao M and Baixaulí C, Librero J as affiliated researchers), University of Southern Denmark's team (Christiansen T, Thygesen LC), University of York's team (Bloor K, Cookson R, Gutacker N), University Nova de Lisboa 's team (Mateus C, Nunes C, Joaquim I), Institute of Public Health of Ljubljana's team (YäzbeK AM, Galsworthy M, Albreht T), UMIT's team (Munck J, Güntert B) and EHMA's team (Bremner J, Giepmans P, Dix O).

Funding

This work was supported by the FP-7 Programme of the European Union with Grant HEALTH-F3-2010-242189 for the 'ECHO' project and complementary funds (grants RD12/0001/0004) from the Instituto de Salud Carlos III subprogram RETICS co-financed by the European Regional Development Fund and the Fundación Instituto de Investigación en Servicios de Salud.

Conflicts of interest: None declared.

Key Points

- Health systems bear substantial opportunity cost in using interventions deemed lower-value. Quantifying the utilization of this type of care and its systematic variation across policy-relevant geographical units could uncover local room for realignment into value-based provision of care.
- C-section and, in particular, its lower-value indication (Caesarean in low obstetric risk deliveries) provide the case study to test this approach in five European countries: Denmark, England, Portugal, Slovenia and Spain.
- The intensity of C-section utilization in lower-value indications ranges between 2 and 10% of all deliveries, alongside large variations across areas within each country. Such behaviour was stable over the period of analysis (2002–9).
- An empirical upper cap of utilization was set in each country and excess-cases calculated per area and region. The analysis conducted suggests plenty of room for enhancing value in obstetric care and equity in women's access to such within ECHO countries.
- The case of lower-value C-section shows how examining geographical variations in lower-value care can be a powerful screening tool for health systems.

References

1 Thomson S, Figueras J, Evetovits T, et al. Economic Crisis, Health Systems and Health in Europe: Impact and Implications for Policy. Policy Summary 12. Copenhagen: WHO Regional Office for Europe and European Observatory on Health Systems and Policies, 2014. Available at: http://www.euro.who.int/en/about-us/partners/observatory/policy-briefs-and-summaries/economic-crisis,-health-systems-and-health-in-europe-impact-and-implications-for-policy (2 January 2015, date last accessed).

2 Garcia-Armesto S, Campillo-Artero C, Bernal-Delgado E. Disinvestment in the age of cost-cutting sound and fury. Tools for the Spanish national health system. *Health Policy* 2013;110:180–85.

3 Elshaug AG, Moss JR, Littlejohns P, et al. Identifying existing health care services that do not provide value for money. *Med J Aust* 2009;190:269–73.

4 Elshaug AG, Hiller JE, Tunis SR, Moss JR. Challenges in Australian policy processes for disinvestment from existing, ineffective health care practices. *Aust New Zealand Health Policy* 2007;4:23.

5 Elshaug AG, Watt AM, Moss JR, Hiller JE. Policy Perspectives on the Obsolescence of Health Technologies in Canada [HTS Policy Forum Discussion Paper]. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH), 2009. Available at: http://www.cadth.ca/media/pdf/Obsolescence%20of%20Health%20Technologies%20in%20Canada_Policy_Forum_e.pdf (2 January 2015, date last accessed).

6 Bohmer P, Pain C, Watt A, et al. Maximising health gain within available resources in the New Zealand public health system. *Health Policy* 2001;55:37–50.

7 Audit Commission. Reducing Spending on Low Clinical Value Treatments*Health Briefing*. London: Audit Commission, 2011. Available at: http://www.rightcare.nhs.uk/index.php/2011/04/audit-commission-health-briefing-reducing-spending-on-low-clinical-value-treatments/ (2 January 2015, date last accessed).

8 NHS Rightcare Commissioning for Value. Available at: http://www.rightcare.nhs.uk/index.php/commissioning-for-value/ (2 January 2015, date last accessed).

9 ECHO Consortium Lower-Value Care Atlas Reports. Available at: http://www.echo-health.eu/ (2 January 2015, date last accessed).

10 Betrán AP, Merialdi M, Lauer JA, et al. Rates of caesarean section: analysis of global, regional and national estimates. *Paediatr Perinat Epidemiol* 2007;21:98–113.

11 Althabe F, Sosa C, Belizán JM, et al. Cesarean section rates and maternal and neonatal mortality in low-, medium-, and high-income countries: an ecological study. *Birth* 2006;33:270–77.

12 Belizán JM, Althabe F, Cafferata ML. Health consequences of the increasing caesarean section rates. *Epidemiology* 2007;18:485–86.

13 Baicker K, Buckles KS, Chandra A. Geographic variation in the appropriate use of cesarean delivery. *Health Aff* 2006;25:w355–67.

14 Librero J, Peiró S, Belda A, Calabuig J. C-section rate in low-risk women: a useful indicator to compare hospitals attending deliveries with different risks. *Rev Esp Salud Pública* 2014;88:315–26.

15 Lauer JA, Betrán AP, Merialdi M, Wojdyla D. Determinants of Caesarean Section Rates in Developed Countries: Supply, Demand and Opportunities for Control. World Health Report, 2010. Background Paper, 29. Geneva: World Heath Organization, 2010. Available at http://www.who.int/healthsystems/topics/financing/healthreport/29DeterminantsC-section.pdf (2 January 2015, date last accessed).

16 Gibbons L, Belizán JM, Lauer JA, et al. The Global Numbers and Costs of Additionally Needed and Unnecessary Caesarean Sections Performed Per Year: Overuse as a Barrier to Universal Coverage. World Health Report, 2010. Background Paper, 30. Geneva: World Heath Organization, 2010. Available at http://www.who.int/healthsystems/topics/financing/healthreport/30C-sectioncosts.pdf (2 January 2015, date last accessed).

17 OECD. Caesarean sections. In: Health at a Glance 2011, OECD Indicators. OECD Publishing, 2011. Available at http://dx.doi.org/10.1787/health_glance-2011-37-en (2 January 2015, date last accessed).

18 European Collaboration for Healthcare Optimization (ECHO) www.echo-health.eu. Zaragoza (Spain): Instituto Aragonés de Ciencias de la Salud—Instituto Investigación Sanitaria Aragón; c2011. ECHO Consortium. Handbook on Methodology: ECHO Indicators, 2014 Apr 27. Available at: http://www.echo-health.eu/handbook/getting-indicators.html (15 October 2014, date last accessed).

19 Ibáñez B, Librero J, Bernal-Delgado E, et al. Is there much variation in variation? Revisiting statistics of small area variation in health services research. *BMC Health Serv Res* 2009;9:60.

20 European Collaboration for Healthcare Optimization (ECHO) www.echo-health.eu. Zaragoza (Spain): Instituto Aragonés de Ciencias de la Salud-Instituto Investigación Sanitaria Aragón; c2011. In: Bernal-Delgado E, Thygesen LC, Martínez-Lizaga N, Comendeiro M, editors. On Behalf of the ECHO Consortium. Handbook on Methodology: Measurement of the Variation, 2014 Apr 27. Available at: http://www.echo-health.eu/handbook/measuring_variation.html (15 October 2014, date last accessed).

21 Kozhimannil KB, Hung P, Prasad S, et al. Rural-urban differences in obstetric care, 2002-2010, and implications for the future. *Med Care* 2014;52:4–9.

Hospital characteristics and Avoidable C-sections: a decade analysis (Work 3)

Inês Joaquim
Céu Mateus
Luís Nobre Pereira
Carla Nunes

Submitted to an international journal

Hospital characteristics and avoidable C-sections: a decade analysis

Inês Joaquim[1], Céu Mateus[2], Luís N. Pereira[3], Carla Nunes[4]

[1] Escola Nacional de Saúde Pública, Universidade Nova de Lisboa

[2] Division of Health Research, Lancaster University

[3] Research Centre for Spatial and Organizational Dynamics (CIEO) - University of Algarve, School of Management, Hospitality and Tourism

[4] Centro de Investigação em Saúde Pública, Escola Nacional de Saúde Pública, Universidade Nova de Lisboa

Abstract

Background: C-sections (CS) have been under the scrutiny of policy makers who intend to reduce the number of these procedures. Our objective is to contribute to the understanding of the role played by hospital characteristics and resources in C-sections that could have been avoided.

Methods: Avoidable C-Section Rates are computed using inpatient data from Portuguese National Health Service hospitals (2002-2011). Fractional response models are estimated to understand how A-CSR are related to hospital characteristics: hospital size, hospital specialization in obstetrics, obstetric resources and teaching status.

Results: The A-CSR increased from 13.6% (2002) to 17.6% (2008) declining to 16.0% in 2011. Bigger hospitals and teaching hospitals present lower A-CSR. The availability of a Neonatal Intensive Care Unit is positively related to A-CSR. The ratio of obstetricians to obstetric beds is not significant. Results show that much of the variation in the A-CSR remains to be explained after controlling for these four characteristics.

Conclusion: Hospital resources are relevant but not sufficient in explaining the variation in A-CSR between hospitals. Further studies are needed considering medical practice, economic incentives and health education to the general population.

Introduction

Caesarean sections (CS) are the most frequently performed surgical procedures in the world [1,2]. In 1985, the World Health Organization (WHO) delivered general recommendations on CS, stating that the minimum acceptable rate is 1% and that there is no justification for any region to have rates higher than 10–15%. In 2015, the WHO once again stated that CSR above 10% are not associated with reductions in maternal and new-born mortality rates [3]. Plus, the negative impact of exposing the mother to unnecessary risks without additional benefit should only be undertaken when medically necessary [2,4].

Nowadays, this level is exceeded in most developed countries, where CS has been rapidly rising to an average of 21% of all births [5]. Around the world, 25.7% of births are delivered by CS, although rates vary significantly across countries suggesting that a number of CS are done without medical need [4,6].

In Portugal, CSR have been increasing and thus have been under the scrutiny of policy makers in order to reduce them. In 2014, after a decade of rising rates, Portuguese NHS hospitals registered a decrease in the rate to a level below 30% [7].

The increase in risk factors (RF) of mothers is the most straightforward justification for the increasing trend in CS procedures. Assuming that, the upper limit recommended by the WHO would be invalid today as too low [8]. This statement generates some controversy since not much evidence exists proving that higher rates provide additional benefits and the increase in CSR may be a result of other situations where clinical need is not present [8].

Other non-clinical justifications for the increasing trends of CSR are presented throughout the literature and include obstetricians' behaviour and medical practice [9–12], provider's financial incentives, as well as the preferences or other time-related incentives for mothers requesting elective CS [4,13–16].

Regarding obstetrician incentives to choose CS over vaginal delivery, the most visible incentive might be the economic one. Evidence shows that with the right financial incentives – similar payments for CS and vaginal delivery, for example – CSR might decrease [7].

Opportunity costs of prolonged deliveries, minimization of malpractice risks or defensive medicine and the demand for leisure or normal workweeks are also identified reasons for the increasing numbers of CS [17].

As for medical practice and resource availability, CSR have been adjusted for RF and pregnancy characteristics and variations between hospitals have been analysed to assess the relationship between rates and hospital characteristics. Differences in hospital CSR can be attributed to pregnancy characteristics, although the increasing incidence of clinical indications for CS along the years do not justify completely the rising rates [18]. Broad variation can be found in preterm CS births [19] but differences in CSR can also be a result of almost random decision making [20]. Variation can also be linked with the uncertainty of a diagnosis and the introduction of diagnostic technology may reduce hospital variation

[21].

The availability of resources motivates the increase in CSR. CSR increased for admissions that occurred from Monday to Thursday and decreased with admissions on Friday and Saturday, when fewer resources were available [4]. In contrast, CS deliveries also increase resources used because these procedures are associated with a length of hospital stay that is two to three days longer compared to a vaginal delivery, higher hospital costs and physician fees [17].

At an aggregate level, CSR are related to income, but supply factors also play a critical role: the greater the capacity of the health system, the greater the number of CS that are performed [22]. Moreover, obstetricians have a substantial influence on the delivery mode [22]. At the same time, the supply characteristics also seem to affect the CSR which decrease when more specialized resources are available, including facilities with a neonatal intensive care unit (ICU) and maternal foetal medicine subspecialists [23]. The rates also decrease with the availability of obstetricians–gynaecologists as opposed to family physicians only, a higher delivery volume, an urban location and 24-hour in-house anaesthesiology. No differences were found between types of hospital ownership nor between teaching status [4,24,25].

Considerable work has been made in relating CSR with available resources, but no work has been found on specific populations, such as low-risk deliveries, that eliminate biases related to hospitals handling with more pregnancies with indication for CS. Looking at CS that were performed without an extended list of medical reasons and are thus considered avoidable, this work aims to understand how hospital characteristics contribute to avoidable C-section rates (A-CSR) and help policy makers designing policies to reduce them.

Methods

Considering that there are CS that should not be performed, this work focuses on avoidable CS (A-CS) which are defined as a CS that was performed although no RF was identified (Table 1). Relevant RF were identified from the literature on the topic and current practice [2,4,9,19,26–28]. Previous CS is excluded from this list as according to current guidelines a vaginal delivery should be planned after a CS [29]. These cases represented 6% of deliveries without RF.

Data combines anonymized data on discharges that took place in Portuguese NHS hospitals between 2002 and 2011 (a decade) collected by the Central Administration of the Health System (Administração Central do Sistema de Saúde) and a dataset on the resources available at each NHS hospital for the same period collected by the Directorate General of Health (Direcção Geral de Saúde). In this analysis, CS and vaginal deliveries grouped in DRGs 370–373 (AP-DRG 21) are included. We exclude high risk deliveries and deliveries with other surgical procedures. For each delivery, an array of additional information was available identifying the hospital where the delivery took place, co-morbidities and other clinical conditions.

Table 1: Risk factors – List of pregnancy characteristics and maternal medical conditions

A-CSR, our dependent variable, is a proportion, therefore linear regression models are not adequate as they may lead to nonsensical predictions (out of bounds of the standard unit interval) and the effects of explanatory variables tend to be non-linear (not constant through all of its range as the variance tends to decrease when the mean gets closer to the boundaries) [30,31]. Fractional Response Models (FRM) shall be used to accommodate to these constraints although they require the correct specification of the conditional expectation of the fractional response variable i.e. a functional form for the distribution of the A-CSR must be assumed. This functional form imposes the constraints on the conditional mean of this variable. Frequent choices for the functional form include logistic, probit, loglog, cloglog and cauchit functions, which were estimated in this study using the Bernoulli-based quasi-maximum likelihood method. Details about these functions may be found in Ramalho et al. [31].

In order to determine if the functional form is correctly specified RESET-type tests and the goodness-of-functional form (GGOFF) tests were performed. Both test if the functional form assumed corresponds to the conditional mean and can be interpreted as tests for the omission of other explanatory variables in the model. If the functional form is correctly specified, the null hypothesis shall not be rejected.

Four hospital characteristics are considered to understand how hospital resources affect A-CSR: hospital size (100s of beds per hospital), hospital specialization in obstetrics (availability of neonatal ICU, binary variable: yes/no), availability of obstetric resources (ratio of obstetricians to obstetric beds) and teaching status (binary variable: yes/no). While availability of resources is expected to increase A-CSR, the specialization of these resources is expected to reduce unneeded care thus reducing this rate.

Understanding how these variables influence A-CSR qualify policy makers to define the optimal level of resources towards A-CSR minimization.

Summary statistics can be found in Table 2.

Table 2: Summary statistics of hospitals' resources

The reduction in the number of hospitals between 2002 and 2011 is due to hospital mergers taking place over the period. This organizational fact influences not only the number of institutions analysed in each year but also the resources of each hospital which were combined as a new institution was created. As consequence, the average number of beds available at hospitals increases over the years as well as the percentage of hospitals with a neonatal ICU and the number of teaching hospitals. No more than three hospitals nationwide have no teaching activities, nevertheless, it was decided that the variable needed to be controlled for to understand whether or not the differences appeared to be systematic.

SPSS was used for the summary statistics, while the Stata FRM package, developed by Ramalho, et al. [31], was used to estimate econometric models.

Results

Deliveries have been constantly decreasing until 2010 when a slight increase occurred (Table 3). CSR steadily increased between 2002 and 2009 from 26% to 31%, at which point the rate seems to start to decrease reaching 29% in 2011. These values are still much higher than the widely quoted 15% threshold recommended by the WHO.

Table 3: Summary statistics on Portuguese NHS deliveries and hospitals

The percentage of deliveries without RF decreased from 69% in 2002 to 60% in 2011, justifying the increase observed in CS as more deliveries register RF. A-CSR has steadily increased between 2002 and 2008, with a slight reversal of the trend from 2009 onwards.

There is a significant dispersion of the A-CSR with values ranging from 5% to as much as 40% (Coefficient of Variation around 38%).

Results from the FRM estimations for the different functional forms are presented in Table 4. For each functional form, two models were estimated, one including only the hospital characteristics and the other additionally including the variable year.

Table 4: FRM estimates for different functional forms (N=442)

The fact that the dependent variable is not truly centred but asymmetric towards 0 would suggest the loglog model as the best choice for the functional form. In fact, the loglog model appears to be the most correct functional form as it is the only functional form that is not rejected by both RESET and GGOFF tests.

In these models, the number of beds and the A-CSR are negatively related meaning that when controlling for other variables, bigger hospitals present lower A-CSR than do their counterparts.

As expected, teaching hospitals are associated with lower levels of A-CSR. In contrast, availability of a neonatal ICU is related to higher values of A-CSR.

The ratio of obstetricians to obstetric beds presents a negative correlation with the A-CSR but it is not significant at a 5% significance level, when the other variables are included in the model.

The coefficient of the variable year is positive and statistically significant. This means that regardless of hospital characteristics, the A-CSR has increased over the years.

For all the models estimated, R^2 is small and thus, selected hospital characteristics only explain a small part of the variations observed in the A-CSR of hospitals.

Discussion

Looking at CS that could have been avoided, we present a novel approach in the literature of CS. When comparing hospitals that handle with different patient's complexity this is especially relevant. With this approach we guarantee that hospitals specialized in the handling of complex cases and which tend to have higher CSR, are not penalized.

So far, the literature has justified the variability in A-CSR with differences in medical practices, behaviours, preferences, hospitals characteristics and availability of resources [9–23]. Availability of resources has been suggested to increase activity in line with the theory that when resources are available they tend to increase activity regardless of its need. When these resources are more specialized, a wiser utilization is reported, though.

Our results should be interpreted with caution: while bigger hospitals, regardless of their focus on obstetric care have more resources, they may be working closer to maximum capacity, promoting wiser utilization of resources. Those same hospitals tend to have more specialized resources, not only in obstetrics but in general, though. The positive correlation between neonatal ICU care and A-CSR suggests that availability of resources is more relevant to explain the behaviour/trend observed in the A-CSR than the potential specialization of resources. Furthermore, our results support teaching status being negatively related to A-CSR, a finding with little or no evidence in the literature so far. Nevertheless, this must be interpreted with caution as 98% of the hospitals in the dataset are teaching hospitals.

When other variables were considered in the models, the ratio of obstetricians to obstetric beds was not significant at a 5% significance level suggesting that the relationship between obstetric resources available is not explanatory of differences in A-CSR.

Including a time variable does not affect the results found for the hospital characteristics considered and indicated that A-CSR is increasing as time goes by.

There are some discrepancies between the results found in this work and the results from other authors. This work uses A-CSR as opposed to using the general CSR or the common nulliparous, term, singleton and vertex births (NTSV) and thus results are not truly comparable. Nevertheless, CSR do not account for CS that are clinical justifiable and NTSV definition has proven not to be sufficient to eliminate potential clinical justification for CS. This lack of comparison also happens when using other metrics and comparing results across countries since there are international discrepancies in the classification of deliveries without complications [32].

Moreover, different models are used in the literature with some authors modelling the probability of CS delivery and other authors modelling hospital variations using fixed or random-effects models. Having a more comprehensive list of RFs allowing the exclusion of the deliveries that had justification for CS was desirable given the database available to work and the choice of FRM to model the investigation question seemed the most appropriate given the fractionary dependent variable being studied.

The small explanatory power of the models reveals that much of the variation in the A-CSR is still to be explained; this includes medical practice for example, as well as other organizational arrangements and mothers' preferences.

Doctors' behaviour and medical practice were pointed in the literature as being one of the main predictors for CS. Medical practices, such as 24-hour in-house anaesthesiologists, high-quality diagnostic technology that standardizes practice style, the practice of obtaining second opinions, training doctors on CS guidelines and auditing have been documented as highly influencing CSR, although there is no consensus. Those variables were not available for analysis.

As for the mothers' preferences, while some preferences may be related to misinformation or socioeconomic status, they rely heavily on the opinion of the doctor [33,34]. This means that somehow, mother's preferences or beliefs may influence doctors' decisions on the mode of delivery although only clinical factors are stated as being indicative for a CS. This work only studied deliveries in public hospitals, excluding biases from different protocols between public and private hospitals. Information on socioeconomic status or maternal request for CS was not available.

Another important influencing factor not tested was the one of economic incentives to the physician, which is frequently put forth as an important factor in decreasing CSR. At the time of the analysis, economic incentives had not been implemented in Portugal therefore their impact has not been assessed.

Additional limitations relate to aggregated information on merged hospitals which made us lose additional insights on individual hospitals, data unavailability of other potential explanatory variables such as anaesthesiologists who are required to perform CS, NHS scope and the role of financial incentives on performance.

Further analysis may be developed using multilevel analysis combining individual characteristics of the mothers and hospital characteristics.

Recommendations, such as auditing and feedback on all CS performed, second opinions, training of doctors on caesarean delivery guidelines and implementation of health education and behaviour change strategies, have been suggested to reduce CSR without affecting the outcomes. Results of these practices are not consensual. Some authors have found positive results [35–37]. Other authors have found modest results [38–40]. The findings of Epstein and Nicholson (2009), for example, suggest that even learning from experience sharing, physicians do not really converge their styles, nor do they revise their prior beliefs. Other suggestions for reducing rates include annual publication of the CSR per hospital and the inpatient rate due to hypoxic-ischemic new-borns, paying the same for a CS or for a vaginal delivery, financial incentives to hospitals that present lower CSR, implementation of an operating theatre next to a delivery room and the implementation of non-induction of labour with no medical reason before 41 weeks gestation [26]. In Portugal, the implementation of such measures seems to be reaching the objectives, as NHS hospitals registered a decrease in the CSR to a level below 30% in 2014 [7].

Conclusion

Our results support the idea that policies on reduction of CS, and namely on A-CS, should not focus on increasing efficiencies by reducing resources, but instead on changing practices. This can be achieved not only by changing economic incentives, as done recently in Portugal, but also by investing in continuous training of doctors, performing peer reviews, obtaining second opinions on CS and auditing all CS cases, as suggested by different authors. Public health campaigns on the risks associated with CS to the general population should also be developed.

References

[1] Althabe F, Belizán J. Caesarean section: the paradox. Lancet [Internet]. 2006 Oct;368(9546):1472–3. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140673606696165

[2] Gibbons L, Belizán J, Lauer J, Betrán A, Merialdi M, Althabe F. Inequities in the use of cesarean section deliveries in the world. Am J Obstet Gynecol. 2012 Apr;206(4):331.e1–19.

[3] World Health Organization. WHO statement on Caeserean Section Rates. 2015.

[4] Aelvoet W, Windey F, Molenberghs G, Verstraelen H, Van Reempts P, Foidart J-M. Screening for inter-hospital differences in cesarean section rates in low-risk deliveries using administrative data: an initiative to improve the quality of care. BMC Health Serv Res [Internet]. 2008;8:3. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2266728&tool=pmcentrez&rendertype=abstract

[5] Betrán AP, Merialdi M, Lauer JA, Bing-Shun W, Thomas J, Van Look P, et al. Rates of caesarean section: Analysis of global, regional and national estimates. Paediatr Perinat Epidemiol. 2007;21(2):98–113.

[6] World Health Organization. Caeserean section without medical indication increases risk of short-term adverse outcomes for mothers. 2010.

[7] Ayres-De-Campos D, Cruz J, Medeiros-Borges C, Costa-Santos C, Vicente L. Lowered national cesarean section rates after a concerted action. Acta Obstet Gynecol Scand. 2015;94(4):391–8.

[8] Gibbons L, Belizán J, Lauer J. The global numbers and costs of additionally needed and unnecessary caesarean sections performed per year: overuse as a barrier to universal coverage. World Heal Rep. 2010;

[9] Bailit JL, Love TE, Mercer B. Rising cesarean rates: Are patients sicker? Am J Obstet Gynecol. 2004;191(3):800–3.

[10] Finsen V, Storeheier AH, Aasland OG. Cesarean section: Norwegian women do as obstetricians do - Not as obstetricians say. Birth. 2008;35(2):117–20.

[11] Habiba M, Kaminski M, Da Frè M, Marsal K, Bleker O, Librero J, et al. Caesarean section on request: a comparison of obstetricians' attitudes in eight European countries. BJOG. 2006 Jun;113(6):647–56.

[12] Lehmann S, Børdahl PE, Rasmussen SA, Irgens LM. Norwegian midwives and doctors have increased cesarean section rates. Acta Obstet Gynecol Scand. 2007 Jan;86(9):1087–9.

[13] Baicker K, Buckles K, Chandra A. Geographic variation in the appropriate use of cesarean delivery. Health Aff. 2006;25(5).

[14] Ma K-ZM, Norton EC, Lee S-YD. Declining fertility and the use of cesarean delivery: evidence from a population-based study in Taiwan. Health Serv Res. 2010 Oct;45(5 Pt 1):1360–75.

[15] McCourt C, Weaver J, Statham H, Beake S, Gamble J, Creedy DK. Elective cesarean section and decision making: A critical review of the literature. Vol. 34, Birth. 2007. p. 65–79.

[16] Schulkind L, Shapiro TM. What a difference a day makes: Quantifying the effects of birth timing manipulation on infant health. J Health Econ [Internet]. 2014 Jan;33:139–58. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0167629613001458

[17] Brown HS. Physician demand for leisure: Implications for cesarean section rates. J Health Econ. 1996;15(2):233–42.

[18] Kozhimannil KB, Law MR, Virnig BA. Cesarean delivery rates vary tenfold among US hospitals; Reducing variation may address quality and cost issues. Health Aff. 2013;32(3):527–35.

[19] Zeitlin J, Di Lallo D, Blondel B, Weber T, Schmidt S, K??nzel W, et al. Variability in caesarean section rates for very preterm births at 28-31 weeks of gestation in 10 European regions: Results of the MOSAIC project. Eur J Obstet Gynecol Reprod Biol. 2010;149(2):147–52.

[20] Clark SL, Belfort MA, Hankins GD V, Meyers JA, Houser FM. Variation in the rates of operative delivery in the United States. Am J Obstet Gynecol. 2007;196(6).

[21] Grytten J, Monkerud L, Sorensen R. Adoption of Diagnostic Technology and Variation in Caesarean Section Rates: A Test of the Practice Style Hypothesis in Norway. Health Serv Res. 2012;47(6):2169–89.

[22] Lauer J a, Betrán AP, Merialdi M, Wojdyla D. Determinants of caesarean section rates in developed countries : supply , demand and opportunities for control. World Heal Organ. 2010;22.

[23] Clark SL, Xu W, Porter TF, Love D. Institutional influences on the primary cesarean section rate in Utah, 1992 to 1995. Am J Obstet Gynecol. 1998 Oct;179(4):841–5.

[24] Snyder CC, Wolfe KB, Loftin RW, Tabbah S, Lewis DF, Defranco EA. The influence of hospital type on induction of labor and mode of delivery. Am J Obstet Gynecol. 2011;205(4).

[25] Han K-T, Kim SJ, Ju YJ, Choi JW, Park E-C. Do hospital characteristics influence Cesarean delivery? Analysis of National Health Insurance claim data. Eur J Public Health [Internet]. 2017;27(5):801–7. Available from: http://academic.oup.com/eurpub/article/27/5/801/3803433/Do-hospital-characteristics-influence-Cesarean

[26] Ayres-De-Campos D, Furtado JM, Crisóstomo MR, Carrapato R, Cunha E, Conceição MF da. Medidas para reduzir a taxa de cesarianas na região Norte de Portugal (Measures to reduce cesarean rate in the North region of Portugal). 2010.

[27] Bragg F, Cromwell D a, Edozien LC, Gurol-Urganci I, Mahmood T a, Templeton A, et al. Variation in rates of caesarean section among English NHS trusts after accounting for maternal and clinical risk: cross sectional study. BMJ. 2010;341:c5065.

[28] Direcção Geral de Saúde. Registo de Indicações de Cesariana (Registry of cesarean indications) [Internet]. 2015. Available from: https://www.dgs.pt/em-destaque/norma-e-orientacoes-sobre-aspetos-tecnicos-relacionados-com-a-cesariana.aspx

[29] Royal College of Obstetricians and Gynaecologists. Birth after previous caesarean birth. R Coll Obstet Gynaecol. 2015;(45):1–17.

[30] Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401 (K) plan participation rates. J Appl Econom [Internet]. 1996;11(6):619–32. Available from: http://www.jstor.org/stable/2285155?origin=JSTOR-pdf

[31] Ramalho EA, Ramalho JJS, Murteira JMR. Alternative estimating and testing empirical strategies for fractional regression models. J Econ Surv. 2011;25(1):19–68.

[32] Or Z, Renaud T, Thuilliez J, Lebreton C. Diagnosis related groups and variations in resource use for child delivery across 10 European countries. Heal Econ (United Kingdom). 2012;21(SUPPL. 2):55–65.

[33] Copelli FH da S, Rocha L, Zampieri M de FM, Gregório VRP, Custódio ZA de O. Determinants of women's preference for cesarean section. Texto Context - Enferm [Internet]. 2015 Jun;24(2):336–43. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-07072015000200336&lng=en&nrm=iso&tlng=en

[34] Teixeira C, Silva S, Severo M, Barros H. Socioeconomic position early in adolescence and mode of delivery later in life: Findings from a Portuguese birth cohort. PLoS One. 2015;10(3).

[35] Althabe F, Belizán J, Villar J, Alexander S, Bergel E, Ramos S, et al. Mandatory second opinion to reduce rates of unnecessary caesarean sections in Latin America: A cluster randomised controlled trial. Lancet. 2004;363(9425):1934–40.

[36] Poma PA. Effect of departmental policies on cesarean delivery rates: A community hospital experience. Obstet Gynecol. 1998;91(6):1013–8.

[37] Socol ML, Garcia PM, Peaceman AM, Dooley SL. Reducing cesarean births at a primarily private university hospital. Am J Obstet Gynecol. 1993;168(6 PART 1):1748–58.

[38] Epstein AJ, Nicholson S. The formation and evolution of physician treatment styles: An application to cesarean sections. J Health Econ. 2009;28(6):1126–40.

[39] Bickell N, Zdeb M, Applegate M, Roohan P, Sui A. Effect of external peer review on cesarean delivery rates: a statewide program. Obstet Gynecol. 1996 May;87(5 Pt 1):664–7.

[40] Frigoletto FD, Lieberman E, Lang JM, Cohen A, Barss V, Ringer S, et al. A clinical trial of active management of labor. N Engl J Med [Internet]. 1995;333(12):745–50. Available from: http://www.nejm.org/doi/full/10.1056/NEJM199509213331201\nhttp://www.ncbi.nlm.nih.gov/pubmed/7643880

Table 1: Risk factors – List of pregnancy characteristics and maternal medical conditions

Maternal age	Under 18 years-old or above 40 years-old
Medical Conditions	**Diagnosis codes (ICD-9-CM)**
Early labour	644.21
Late labour	645.21
Multiple gestation	651.x1
Mal presentation of foetus	652.x1
Disproportion	653.x1
Maternal pyrexia	659.21
Septicaemia	659.31
Complications of pregnancy or labour	640.x1
	641.x1
	642.x1
	643.11, 643.21, 643.81
	646.x1
	647.x1
	648.x1
	660.x1

Table 2: Summary statistics of hospitals' resources

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Number of hospitals	50	50	50	50	49	40	40	40	40	39
Hospital dimension (100s of beds per hospital)										
Average	381	381	381	382	398	505	505	508	508	592
Standard deviation	274	278	279	277	281	299	305	304	297	382
Minimum	97	97	97	97	97	143	143	138	139	126
Maximum	1,525	1,530	1,548	1,505	1,497	1,496	1,456	1,456	1,375	2,070
Hospital specialization in obstetrics (availability of neonatal ICU: yes/no)										
% of hospitals (yes)	28%	32%	32%	30%	37%	40%	40%	45%	45%	36%
Availability of obstetric resources (ratio of obstetricians to obstetric beds)										
Average	0.363	0.344	0.347	0.333	0.337	0.359	0.371	0.364	0.388	0.375
Standard deviation	0.144	0.120	0.123	0.118	0.122	0.130	0.149	0.147	0.200	0.160
Minimum	0.105	0.154	0.125	0.000	0.053	0.068	0.125	0.098	0.125	0.125
Maximum	0.857	0.857	0.886	0.686	0.657	0.686	0.755	0.819	1.191	0.864
Teaching status (teaching hospital: yes/no)										
% of hospitals (yes)	92%	94%	94%	98%	98%	98%	98%	98%	95%	95%

Source: *Direcção Geral de Saúde*

Table 3: Summary statistics on Portuguese NHS deliveries and hospitals

Year	N	Total Deliveries	CSR	% of Deliveries Without Risk Factors out of Total Deliveries	National A-CSR	A-CSR at hospital level		C.V. μ/σ
						Mean (S.D.)	Min – Max	
2002	50	93,649	26.0	69.4	13.6	14.1 (5.3)	4.9 – 27.4	0.377
2003	50	90,888	27.6	67.8	14.5	15.4 (5.6)	7.1 – 27.2	0.363
2004	50	88,086	28.5	66.3	14.6	15.4 (5.7)	5.7 – 28.5	0.371
2005	50	87,701	30.1	65.8	16.1	17.1 (6.6)	7.4 – 36.8	0.387
2006	49	85,023	30.5	65.4	16.5	17.8 (7.0)	5.8 – 40.3	0.394
2007	40	82,309	31.0	65.2	17.3	17.2 (6.9)	7.3 – 28.6	0.401
2008	40	82,696	31.6	64.2	17.6	17.4 (6.7)	5.4 – 30.6	0.383
2009	40	78,752	31.6	62.4	17.3	17.3 (6.6)	6.2 – 31.5	0.383
2010	40	71,058	30.3	61.6	16.7	16.6 (6.0)	6.3 – 34.1	0.362
2011	39	73,963	29.1	59.8	16.0	16.4 (5.8)	6.3 – 31.2	0.354

CSR – Caesarean Section Rate; A-CSR – Avoidable Caesarean Section Rate; S.D. – standard deviation; C.V. – coefficient of variation

Table 4: FRM estimates for different functional forms (N=442)

A-CSR (A-CS per hundred CS)	**Logit**		**Probit**		**LogLog**		**CLogLog**		**Cauchit**	
Constant	-1.238**	-49.067*	-0.758**	-27.334*	-0.403**	-22.621*	-1.367**	-45.050**	-1.144**	-83.355**
Number of beds (100 beds)	-0.019*	-0.024**	-0.010*	-0.013**	-0.008*	-0.010**	-0.018*	-0.022**	-0.040*	-0.049*
Availability of Neonatal ICU	0.148*	0.148*	0.082*	0.081*	0.068*	0.067*	0.136*	0.137*	0.269*	0.278*
Obstetricians per obstetric beds	-0.044	-0.059	-0.024	-0.034	-0.020	-0.030	-0.041	-0.052	-0.084	-0.070
Teaching hospital	-0.366**	-0.368**	-0.208**	-0.209**	-0.179**	-0.179**	-0.329**	-0.332**	-0.554**	-0.565**
Year		0.024*		0.013*		0.011*		0.022*		0.041**
R^2	0.0593	0.0810	0.0591	0.0806	0.0589	0.0802	0.0594	0.0812	0.0609	0.0833
Reset test										
LM(2)	4.159*	2.372	3.873*	2.380	3.566	2.411	4.343*	2.358	7.110*	2.377
LM(3)	11.844*	3.150	11.628*	3.082	11.363*	2.988	11.985*	3.216	13.609*	4.214
Goodness of Functional Form (GGOFF) test										
GOFF1 (LM)	3.826	2.293	3.583	2.298					5.682*	1.968
GOFF2 (LM)	3.184	2.147	3.922*	2.394	3.612	2.418	3.785	2.232	3.980*	1.513
GGOFF (LM)	11.738*	3.079	11.138*	3.565	3.612	2.418	3.785	2.232	13.283*	3.808

A-CSR – Avoidable Caesarean Section Rate; A-CS – Avoidable Caesarean Section; CS – Caesarean Section; LM – Lagrange Multiplier

*p-value < 0,05; **p-value<0,001

What have we learnt on measuring hospital efficiency when stochastic frontier analysis is used – Lessons from previous studies (Work 4)

Inês Joaquim
Céu Mateus
Carla Nunes

Submitted to an international journal

What have we learnt on measuring hospital efficiency when stochastic frontier analysis is used – Lessons from previous studies

Inês Joaquim[1], Céu Mateus[2], Carla Nunes[3]

[1] Escola Nacional de Saúde Pública, Universidade Nova de Lisboa, Portugal

[2] Division of Health Research, Lancaster University, UK

[3] Centro de Investigação em Saúde Pública, Escola Nacional de Saúde Pública, Universidade Nova de Lisboa, Portugal

Abstract

Introduction: Financial constraints faced by health care systems have kept efficiency as a central topic for research. Accordingly, efficiency measurement literature has proliferated. Currently, there are two main methods to estimate efficiency – data envelopment analysis (DEA) and stochastic frontier analysis (SFA). SFA method is more demanding and its use is scarce in comparison to DEA. Nevertheless, its parametric approach has advantages by allowing the inclusion of an error component that represents the random shocks to which hospitals might be subject to.

Objectives: To identify the main findings about efficiency assessment using SFA by discussing how hospital resources and characteristics affect institutions' efficiency and what are the main limitations of using SFA.

Methods: Previous literature reviews on the topic were used to collect studies on hospital efficiency measurement. Pubmed and Science Direct were screened to identify other works published recently. Works on hospital efficiency measurement were analysed.

Results: Forty-one studies were identified. Three of them estimated the optimal production function while the others estimated the cost-function. The Cobb-Douglas and the translog model were the most commonly used functional forms for the production function. Almost all the studies decided to test which error distribution fitted the best. Variables selected to perform the analysis were similar across studies and can be divided in output variables, variables of labor and capital resources and hospital characteristics. Results on efficiency levels cannot be compared across studies given differences in the vectors of variables selected to perform the analysis. Results on how variables affect efficiency are not consensual across studies.

Conclusions: Conceptually, efficiency measurement using SFA seems adequate but performing efficiency measurements in the healthcare sector has limitations that were identified across the literature. These limitations may misspecify the model and distort the results. The method may be used but conclusions on efficiency levels and how variables interact with efficiency must be taken with caution.

Introduction

Efficiency has been kept at the center of the research in health care due to its relevance for policy making. However, its measurement has proved to be a challenging topic. Not only outputs produced need to be carefully defined to be comparable, but also inputs, such as work, wages or costs for instance, present challenges in measurement. Moreover, the definition of performance or efficiency can vary according to what must be measured.

Pursuing efficiency is key when resources are limited, and policy-makers and hospital managers struggle with cost-containing policies. Good efficiency measurements are essential not only to assess efficiency levels but also to understand how these levels are affected by market characteristics.

Defining efficiency has not been consensual. Definitions such as the best use of resources in production (Hollingsworth, Dawson, and Maniadakis 1999), the extent to which objectives are achieved in relation to the resources consumed (Jacobs, Smith, and Street 2006) or the production of as many possible outputs given an available set of inputs (Farrell 1957) have been accepted through the literature. Different definitions have not affected what is at its core: institutions that put available resources at use to produce more – more output, better quality, or achieve their objectives easily – are considered more efficient.

The more common methods to estimate efficiency in the healthcare sector are the Data Envelopment Analysis (DEA) – a non-parametric method – and the Stochastic Frontier Analysis (SFA) – a parametric method.

Although efficiency measures using DEA are more common, literature on SFA has been proliferating (Hollingsworth 2003; Hollingsworth 2008).

This paper aims to identify the main findings about efficiency assessment using SFA. We seek to discuss how hospital resources and characteristics affect institutions' efficiency and what are the main limitations of using SFA.

The Stochastic Frontier Analysis

SFA is a parametric method to estimate the optimal production or cost function.

The model was developed simultaneously by Aigner, Lovell, and Schmidt (1977) and Meeusen and van Den Broeck (1977) and is specified as:

$$y_i = f(x_i;\beta) + \varepsilon_i \quad \text{and} \quad \varepsilon_i = u_i + v_i$$

Where y_i represents the output of the hospital and f is the functional form of the technology. The vector of inputs to include in the analysis is represented by x_i and β is the vector of parameters to be estimated. Alternatively, it is possible to estimate a cost-function.

The structure of the error term is the key point of the SFA by being composed by a symmetric disturbance v_i that is independently and identically distributed as $N(0,\sigma_v^2)$ and represents the random fluctuations out of hospital control. It is assumed that u_i and v_i are independent and u_i represents the inefficiencies of the hospital (Aigner, Lovell, and Schmidt 1977; Coelli, Prasada Rao, and Battese 1998).

Technical efficiency is defined as the percentage of the production function attained by the hospital and is the ratio of observed output relative to the potential output of the hospital given its input level (Coelli, Prasada Rao, and Battese 1998):

$$TE_i = \frac{y_i}{exp(x_i\beta)} = \frac{exp(x_i\beta - u_i)}{exp(x_i\beta)} = exp(-u_i)$$

The estimation of the production function via SFA requires the specification of the structure of the production function (Coelli, Prasada Rao, and Battese 1998; Jacobs, Smith, and Street 2006). The more commonly used functions are the Cobb-Douglas and the Translog (Coelli, Prasada Rao, and Battese 1998). The translog function is a generalization of the Cobb-Douglas and thus it has less restrictive properties than the latter. The Cobb-Douglas imposes constant input elasticities, constant returns to scale and the assumption that elasticity of substitution equals one. On the other hand, the Translog is sensitive to multicollinearity and has problems with degrees of freedom (Coelli, Prasada Rao, and Battese 1998). Although, there are differences, authors such as (Zuckerman, Hadley, and Iezzoni 1994), Rosko and Mutter (2008) and Rosko and Mutter (2011) state that the inefficiency estimates are not very sensitive to the choice of the functional form.

SFA requires, as well, that the distribution of the one-sided non-random part of the error term is defined a priori (Coelli, Prasada Rao, and Battese 1998). There are four available distributions (half normal, truncated-normal, exponential and gamma distribution) and there is a consensus across the literature that SFA results are robust to the choice of the error distribution (Zuckerman, Hadley, and Iezzoni 1994; Rosko and Mutter 2008; Rosko and Mutter 2011).

The method in analysis estimates the production function considering only a single output or an index in alternative. In a sector where the outputs produced are, in general, heterogeneous this limitation is substantial. The alternative estimation of a cost-function allows the inclusion of different sets of outputs as separate independent variables (Rosko and Mutter 2011).

Methodology

Studies measuring efficiency have been proliferating along the last decades. To identify previous studies using SFA to measure hospital efficiency, published literature reviews on the topic were used (Hollingsworth, Dawson, and Maniadakis 1999; Hollingsworth 2003; Hollingsworth 2008; Rosko and Mutter 2008).

Pubmed and Science Direct were then screened using key terms such as “stochastic frontier”, health and hospital to identify studies published from 2007 until 2017. Papers about efficiency analysis of hospital services and written in English were selected and analysed.

An additional and final search in Google Scholar was conducted to identify any other work on the subject.

Results

The search returned 43 results. Tables 1-4 present the characteristics of the studies identified – authors, number of institutions considered in the analysis as well as the country and the time period, the selected structure of the production function, the distribution of the error term considered, the variables selected to perform the analysis and the main results.

Most of the studies focus on USA data (26 studies). The remaining studies are mainly focused on European countries: Switzerland and Germany have 3 studies each, Holland, Finland and Spain have 2 studies each and there are also three other studies focusing on England, Portugal and Ireland. Australian hospitals were also studied in one work. Only 1 study uses data from different countries.

There are studies focusing on one year only while others compute efficiency on a panel data basis and others that estimate efficiency for different years. The number of observations also varies from a small sample of 40 hospitals to more than 4.000 hospitals.

Production vs Cost-Function

As said before, SFA analysis deals with a dependent variable that is either the production (for example number of discharges, number of outpatient visits, etc.) if one is estimating the production function or the costs if it is intended to estimate the cost-function.

The healthcare sector, however, is characterized by having multiple and heterogenous outputs. Estimate the production function through SFA is a limitation as it is a difficult task to homogenize the output or researchers are left with summing outputs that are not similar. Due to this, many authors prefer to estimate a cost-function instead of computing a production function. The cost-function by having costs as the dependent variable allows for the inclusion of different outputs produced by the hospitals as separate independent explanatory variables (Rosko and Mutter 2011).

Of the 43 studies, only four (Gannon 2005; Herr 2008; Herwartz and Strumann 2012; Mateus, Joaquim, and Nunes 2015) estimate production functions. Gannon (2005) opted to aggregate inpatient discharges, outpatient visits and day cases in one single output, while Herr (2008), Herwartz and Strumann (2012) and Mateus, Joaquim, and Nunes (2015) weighted the cases to obtain comparable number of cases. All other studies have a cost-function approach using hospital date representing either total costs, operating costs, variable costs or cost per discharge.

Table 1: Main characteristics of the studies using Cobb-Douglas function

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
Vitaliano and Toren (1996)	219/USA/1991	Cobb-Douglas	Half-normal Exponential	Total costs	Patient days, emergency and outpatient visits, teaching hospitals, wages, CMI, technology index, occupancy rate.	Estimated cost-function: except for occupancy rate, all other variables increase costs. Efficiency level: Average level of 18% inefficiency. Other: Facilities with larger Medicare populations are more efficient; hospitals over 300 beds are more efficient, but excessive bed capacity increase costs.
Carey (2003)	1.209/USA/1998	Cobb-Douglas	Half-normal	Total costs	Adjusted admissions, adjusted patient- days, FTE personnel wages, CMI, risk- adjusted mortality index, teaching.	Hospital consolidations that centralize around physician arrangements and insurance products are more cost-efficient. There are potential efficiency gains from hospital consolidations.
Sari (2003)	125/USA/ 1990-1997	Cobb-Douglas	Half-normal Exponential	Total costs	Admissions, outpatient visits, wages, CMI, teaching, patient characteristics.	Mean hospital inefficiency is 19,5%. Efficiency has an inverted u-shape relationship with market concentration; Higher managed care penetration over time is associated with greater hospital efficiency.
Gannon (2005)	210/Ireland/ 1995-2000	Cobb-Douglas	Truncated	Inpatient discharges, outpatient visits, day cases aggregated in one single output	Beds, medical staff, non-medical staff.	Increase in medical staff and number of beds has a significant effect on productivity in regional and general hospitals. SFA efficiency scores estimated are lower than DEA estimates. Regional/general hospitals are highly efficient while county hospitals are less efficient.
Farsi and Filippini (2006)	747/Switzerland/ 1998-2001	Cobb-Douglas	Half-normal	Total costs	Adjusted admissions, patient-days, outpatient revenues, price of capital, price of labour, nurses per bed, ALOS, emergencies, geriatrics.	With exception of geriatrics all other variables increase total costs. Mean inefficiency was 20%. Differences among various ownership subsidization types are not statistically significant; There are unexploited economies of scale in the sample; Larger hospitals are generally more costly; University hospitals are more costly; Ambulatory care is less costly than inpatient care.
Bernet, Rosko, and Valdmanis (2008)	1.757/USA/2001	Cobb-Douglas	Not available	Total cost	Price of capital, admits, outpatient visits, teaching hospital, CMI, mortality index, high-tech index, emergency room share, % Medicaid, % Medicare, government hospital, for-profit, hospital part of a system, hospital competition, market penetration, Mills ratio.	Mortality index and emergency room share have a negative impact on total cost. All other variables considered have a positive impact. Access to debt is associated with lower Inefficiency; Inefficiency increased in the years following a bond issue.

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
Farsi (2008)	168/Switzerland/ 1998-2003	Cobb-Douglas	Ad-hoc	Total costs	Hospitalizations, CMI, average cost per hospitalization, patient days, outpatient revenues, beds, capital price, price of physicians, price of other employees, medical training position, private-insurance admissions.	All variables influence positively the total costs. Hospitals have adopted measures to reduce hospitalizations and empty beds; There is no evidence that a specific ownership type is more or less efficient than other.
Farsi and Filippini (2008)	148/Switzerland/ 1998-2003	Cobb-Douglas	Half-normal	Total costs	Hospitalizations, CMI, average cost per hospitalization, patient days, outpatient revenues, beds, capital price, price of physicians, price of other employees, nurse per bed, medical service centres, non-medical unit, and accredited training positions.	Non-medical units, ambulatory clinics, emergency room and geriatrics have a negative relationship with total costs. All other variables have a positive relationship. Teaching hospitals and hospitals with a broader range of specialization are relatively more costly; After controlling for teaching status, teaching hospitals have still shown a relatively low cost-efficiency; There is no evidence of significant efficiency differences across ownership/subsidy categories.
Ludwig, Groot, and Van Merode (2009)	118/Holland/ 1996-2003	Cobb-Douglas	Time-invariant error	Total costs	Inpatient, day-care, outpatient, cost per patient, administration, cleaning, laundry, kitchen, maintenance, education, laboratory, pharmacy, physiotherapy, average medical, average non-medical.	Large hospitals tend to outsource less, which is supported by efficiency reasons. For most hospital services, outsourcing does not significantly affect the efficiency of hospitals. The focus on the make-or-buy decision may therefore be less important than often assumed.
Ludwig, Van Merode, and Groot (2010)	108/Holland/ 1998-2002	Cobb-Douglas	Time-invariant error	Total costs	LOS, ward days, day-care, admissions, ambulatory care patients, CMI, cost per patient, cost per average DRG inpatient, quality measures, price of intermediate products.	Efficiency and quality are correlated. There is no relationship between the efficiency of departments and the efficiency of the entire hospital. Efficiency estimates varied between 61% and 84% depending on the department.
Pratt (2010)	270/USA/ 2006-2007	Cobb-Douglas	Not-specified	Total costs	Price of labour, price of capital, admissions, outpatient visits, non-acute days, teaching, CMI, Service Mix Index, heart attack, heart failure and pneumonia death rate, % Medicare admissions, for-profit hospital, government hospital, hospital part of a system of hospitals, hospital competition, positive/negative cash-flow.	Price of capital, non-acute days, Medicare case-mix index, heart failure and pneumonia death rate are negatively correlated to total costs. The findings indicate that hospital free cash-flow is significantly linked to firm efficiency/inefficiency. The results indicate that higher positive cash flows are related to lower cost inefficiency, but higher negative cash flows are related to higher cost inefficiency. Thus, cash flows not only impact the ability of hospitals to meet current liabilities, they are also related to the ability of the hospitals to use resources effectively.
Vitikainen, Linna, and Street (2010)	40/Finland/2005	Cobb-Douglas	Half-normal	Operating costs	Admissions, day-care, outpatient visits, capital, doctor and nurses' wages, teaching, readmissions, surgical specialty, obstetric specialty, paediatric specialty, cancer treatments, lung treatments.	Outpatient services have a smaller impact on total costs than inpatient services. At the same time, increased outpatient activity appears to have an adverse effect on estimated cost-efficiency. This counterintuitive finding is probably due to the low weight given to outpatient activities by the Diagnosis Related Groups (DRG) system.

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
Herr, Hendrik, and Boris (2011)	541/Germany/ 2002- 2006	Cobb-Douglas	Truncated	Total adjusted costs	Cost per bed, cost per case, weighted number of cases, beds, doctors, nurses, other staff, labour costs, Eastern Germany, urban, Herfindahl-Hirshman index, ratio of elderly patient, surgeries, length of stay, CMI, mortality.	Input prices influence positively the cost frontier. There are no significant differences in cost efficiency but higher profit efficiency of private than of publicly owned hospitals.
Mateus, Joaquim, and Nunes (2015)	19/Slovenia/ 2005-2009 56/Portugal/ 2002-2009 161/England/ 2005-2008 241/Spain/ 2003-2009	Cobb-Douglas	Half-normal Truncated-normal Exponential	Weighted hospital discharges	Number of employees, number of physicians, number of nurses, number of beds.	SFA is not statistically different from OLS in Portuguese data, while SFA and OLS estimates are statistically different for England, Spain and Slovenia. Panel data should be preferred over cross-section analysis because results are more robust. The most relevant inputs for the production process are the number of beds and the number of employees. Decision makers and hospital managers should consider that a better management of beds may improve efficiency. Measures that improve occupancy rates and help to reduce waiting lists will have a positive impact on efficiency levels.

FTE – Full-time equivalent; CMI – case-mix index; LOS – Length of stay; DRG – diagnosis-related group.

Table 2: Main characteristics of the studies using Translog function

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
Zuckerman, Hadley, and Iezzoni (1994)	1.600/USA/1987	Translog	Half-normal Exponential	Total costs	Admissions, post-admission days, FTE personnel, measures of illness severity, output quality, patient outcomes, costs.	Efficiency estimates depends on which variables enter the cost-function. Mean inefficiency is 18,8% if only input prices are considered. It decreases to 13% if hospital-level variables are used. Efficiency could be improved by reducing and reorganizing staff and gradually downsizing capital stock of hospitals.
Chirikos (1998)	186/USA/ 1982-1993	Translog	Half-normal Exponential	Total costs	Cost per case, cases, beds, teaching, competitive market share, population per square mile, physicians per 100.000 persons, CMI.	Mean inefficiency estimated is 16%. Government hospitals are more likely to be on top levels of efficiency.

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
Rosko (1999)	3.262/USA/1994	Translog	Half-normal Exponential Truncated	Total costs	Discharges, outpatient visits, teaching status, post-admission days, CMI, percentage emergencies	Costs increase with input prices. Teaching hospitals have also higher costs. For profit hospitals were less efficient than not-for-profit hospitals. Hospitals performing in more competitive environments are more efficient.
Rosko and Chilingerian (1999)	195/USA/1989	Translog	Half-normal Exponential Truncated	Total costs	Inpatient discharges, outpatient visits, wage rate, price of capital, severity of illness index, CMI.	Adding CMI adjustment reduces inefficiency scores by 50%. Inefficiency is inversely associated with regulatory pressures and industry concentration.
Deily, McKay, and Dorner (2000)	790/USA/ 1986-1991	Translog	Half-normal	Total costs	Inpatient days, outpatient visits, FTE personnel wages, beds, percentage ICU beds, percentage emergencies, percentage surgeries, high technology service index.	Mean relative inefficiency was 15%. For profits hospitals had the higher inefficiency levels, followed by government hospitals and not-for profit hospitals.
Folland and Hofler (2001)	2.007/USA/1985	Translog	Not available	Total costs	Medical and surgical inpatient days, paediatric, obstetrical and all other inpatient days, outpatient visits, nursing payroll, FTE personnel wages, beds, percent board certified, reservation quality.	With the exception of general inpatient days which have a negative effect on costs, all other hospital services increase costs. Input prices also increase costs. Ownership, teaching status, urban, percentage of medicare and independent hospital showed no efficiency differences when compared with their counterparts.
McKay, Deily, and Dorner (2002)	4.075/USA/ 1986,1991	Translog	Half-normal	Total costs	Inpatient days, outpatient visits, FTE personnel wages, beds, percentage ICU beds, percentage emergencies, percentage surgeries, high tech service index.	Mean inefficiency was highest for for-profit hospitals and lowest for not-for-profit hospitals, with government hospitals falling in the middle.
Rosko (2004)	616/USA/ 1990-1999	Translog	Half-normal Truncated	Total costs	Adjusted discharges, outpatient visits, teaching status, full-time equivalent residents trained, percentage of outpatient surgery, and percentage of Health Maintenance Organizations (HMO).	Decreases in inefficiency were associated with penetration rate and time. Increases in inefficiency were associated with for-profit ownership status and Medicare share of admissions.
Deily and McKay (2006)	140/USA/ 1999-2001	Translog	Exponential Half-normal	Total costs	Physician per bed, residents per bed, CMI, transplant cases, open heart cases per total cases.	Hospitals in the high-performing group are more likely to be for-profit, to have higher occupancy rates, to have proportionately more Medicare and proportionately fewer Medicaid and self-pay patients, to use fewer patient-care personnel per admission, and to have higher operating margins than all other hospitals. Results indicate a positive and significant relationship between hospital's mortality performance and its inefficiency.

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
Yaisawarng and Burgess (2006)	131/USA/2000	Translog	Exponential	Total costs	Output (discharges) divided in basic and complex care, outpatient care, salaries, beds, ICU score, occupancy rate, waiting days, quality measures, urban, teaching, mental.	Average efficiency level was 94%. Marginal cost for treating complex care patients is far less than the respective national capitation rate.
J. Wang, Zhao, and Mahmood (2006)	114/Australia/ 1997-1998	Translog	Half-normal	Total costs	Average salaries of medical and non- medical labour services, inpatient index service, occasion of services, average available beds, ALOS, cost per outpatient, cost per emergency, emergencies.	Inefficiency accounts for 9,3% of total hospital costs in large hospitals and 11,3% in small hospitals, when including complexity indicators. Diseconomies of scale exist in very large hospitals, whereas scale economies appear in very small hospitals. Scope effects are found in both large and small hospitals. Small hospitals are more labor-intensive than large hospitals.
Carey and Dor (2008)	278/USA/ 1991-1998	Translog	Ad-hoc	Total costs	Adjusted patient days, average length of stay, price of labour, labour, price of capital, beds, CMI, rural location, government control, nonprofit status, multi-hopital member, church operated.	Contract adoption leads to reduced expense preference behavior, but that depends critically on the input being examined. Contract-management firms are able to introduce efficiencies over conventional, salaried managers.
Carey, Burgess, and Young (2008)	1.018/USA/ 1998-2004	Translog	Half-normal	Total costs	Discharges, outpatient visits, ALOS, local area wage rates, CMI, outpatient surgeries, patient safety indicators, Herfindahl-Hirshman index, ownership, multi-hospital member, teaching, beds.	There is no evidence that specialty hospitals are more efficient than full-service hospitals. Orthopedic and surgical specialty hospitals appear to have significantly higher levels of cost inefficiency. Cardiac hospitals, do not appear to be different from full-service hospitals.
Mutter, Rosko, and Wong (2008)	1.290/USA/2001	Translog	Truncated	Total costs	Inpatient admissions, outpatient visits, patient days, high technology services, emergency rate, outpatient surgery rate, teaching, mortality, risk-adjustment measures.	Inefficiencies estimated previously can be attributed to variations inpatients burden of illness. Choices about controlling for quality and patient burden of illness can have a nontrivial impact on mean estimated hospital inefficiency and the relative ranking of hospitals generated by SFA.
Barros, de Menezes, and Vieira (2013)	51/Portugal/ 1997-2008	Translog	Half-normal	Variable costs	Discharges, LOS, external consultations, emergency visits, mix of specialist services, beds, price of labour, regional dummies, organizational status, management system, time trend.	Three statistically significant segments in the sample are identified. The health policy based in the identified segments enables a more accurate and cost-effective management of resources. There is evidence that private hospitals are more efficient than public hospitals. Hospitals serving more than one county perform efficiently.

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
Izón and Pardini (2017)	N.S/California (USA)/2005-2013	Translog	Truncated-normal	Total costs	Labour price, discharges, outpatient visits, beds, case-mix index, outpatient surgical care visits, emergency department visits, performance scores, pre- and post-recession indicator, degree of market concentration, ownership, percent of Medicaid patients and discharges, percent of patients with insurance plan, uninsured patient discharges.	The average level of hospital cost inefficiency increased from 10.06 percent to 14.25 percent during the Great Recession and leveled off at 14.03 percent in succeeding years. Higher hospital costs may be an unintended consequence of improved quality of care. Market competition is associated with higher cost inefficiency. The number of full-time employees has a positive correlation with cost-inefficiency. Safety-net hospitals prevented further attrition of cost-efficiency after the great recession though cost-efficiency remained below the pre-recession years levels.

FTE – Full-time equivalent; CMI – case-mix index; LOS – Length of stay; ALOS – average length of stay; N.S. – Not Specified.

Table 3: Main characteristics of the studies using both Cobb-Douglas and Translog function

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
Rosko (2001a)	1.966/USA/1997	Cobb-Douglas Translog	Half-normal Exponential	Total costs	FTE personnel wages, beds, CMI adjusted discharges, outpatient visits, teaching status, percentage emergencies, percentage outpatient surgery	Average estimated X-inefficiency in study hospitals was 12,96%. Increases in managed care penetration, dependence on Medicare and Medicaid, membership in a multihospital system, and location in areas where competitive pressures and the pool of uncompensated care are greater were associated with less X-inefficiency. Not-for-profit ownership was associated with increased X-inefficiency.
Rosko (2001b)	1.631/USA/ 1990-1996	Cobb-Douglas Translog	Half-normal Exponential	Total costs	FTE personnel wages, beds, discharges, outpatient visits, teaching status, percentage emergencies, percentage outpatient surgery	Mean estimated inefficiency decreased by about 28% during the study period. Inefficiency was negatively associated with health maintenance organization (HMO) penetration and industry concentration. It was positively related with Medicare share and for-profit ownership status.
Rosko and Proenca (2005)	1.368/USA/1998	Cobb-Douglas Translog	Truncated Half-normal	Total costs	Outpatient visits, FTE personnel, beds, teaching, CMI, percentage emergencies, percentage surgeries	Mean hospital X-inefficiency was 14,85%. Hospitals providing a moderate to high proportion of services at the network or system level were more efficient than hospitals that did not use networks or systems for service provision. Low users of networks or systems and nonusers had comparable levels of efficiency.

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
Herr (2008)	1.500/Germany/ 2001-2003	Cobb-Douglas Translog	Truncated	Adjusted costs Weighted cases (2 different analysis)	Cases, weighted cases, beds, doctors, nurses, other staff, labour prices, East Germany, female ratio, elderly ratio, surgery ratio, occupancy rate, nurse per bed, LOS, mortality rate, cost per bed.	Efficiency seems negatively correlated with average length of stay, which is highest in private hospitals. Private and non-profit ownership are associated with both lower cost efficiency and lower technical efficiency when compared with public ownership.
Kumar (2010)	136/USA/2005	Cobb-Douglas Translog	Ad-hoc	Total cost per discharge	Salary per discharge, depreciation and interest cost per discharge, supply cost per discharge, liability cost per discharge.	Results suggest that specialty hospitals are significantly more efficient than general hospitals.
Herwartz and Strumann (2012)	1.643/Germany/ 1995-2006	Cobb-Douglas Translog	Truncated	Weighted cases	Beds, physicians, nurses, non-medical staff, material expenses, market share, specialization index, mortality, occupancy rate, hospital budget per bed, elderly ratio, GDP per capita, population density.	Results suggest an increase in the magnitude of negative spatial spillovers related with the expected rise of competition for low cost patients invoked by the announcement or advent of the new financing system.

FTE – Full-time equivalent; CMI – case-mix index; LOS – Length of stay.

Table 4: Main characteristics of the studies using other functional forms

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
Wagstaff (1989)	49/Spain/ 1977-1981	Feldstein cost function	Half-normal	Average cost per case	6 case-mix categories, beds, teaching status, cases per bed.	Estimated cost-function: beds present a u-shape relationship with cost; teaching increases costs; internal medicine has a negative relationship with costs while the other case-mix categories have a positive relationship. Efficiency level: Hospitals of the sample are performing at full efficiency.
Wagstaff and López (1996)	43/Spain/ 1988-1991	Granneman et al. 1986 function	Time-invariant error	Operating costs	Ambulatory visits, emergency cases, inpatient discharges, teaching status, CMI, scanners, rehabilitation, day hospital, oncology, theatres.	Estimated average inefficiency was 58%. There seems to exist economies of scale for emergency visits but diseconomies of scale for ambulatory visits.
Chirikos (1998a)	186/USA/ 1982-1993	Ad-hoc	Half-normal	Total costs	Output, factor prices, CMI.	Frontier methods do not yield sufficiently unambiguous results to serve the short-term needs of hospital regulators.
Linna (1998)	43/Finland/ 1988- 1994	Not specified	Half-normal Truncated.	Operating costs	Outpatient visits, inpatient discharges, inpatient days, bed days, teaching, readmission	Average inefficiency estimated was 7-9%. The readmission rate was found to explain cost-efficiency scores in the parametric models.

Authors	N/Country/Period	Functional Form	Error distribution	Dependent variable	Independent variables	Results
					rate, nurses, scientific publications, price index.	Another possible source of inefficiency is the variation in quality of care which affects the efficiency.
Chirikos and Sear (2000)	186/USA/ 1982-1993	Ad-hoc	Half-normal	Total costs	Cases, CMI, cost per case, FTE per 1.000 cases, LOS, beds, occupancy rate and teaching status.	Mean inefficiency is about 15%. DEA and SFR models yield convergent evidence about hospital efficiency at the industry level, but divergent portraits of the individual characteristics of the most and least efficient facilities.
Li and Rosenman (2001)	90/USA/ 1988-1993	Leontief Translog	N/A	Total costs	Patient days, outpatient visits, labour input prices, capital and other input prices, CMI, Western region, urban, for- profit, Medicare rates.	Hospitals with higher case mix indices or more beds are less efficient while for-profit hospitals and those with higher proportion of Medicare patient days are more efficient. Average efficiency is around 67%.
Street (2003)	226/England/1999	Ad-hoc	Half-normal; Exponential.	Total costs	Inpatient discharges, outpatient attendances, emergency attendances, transfers in and out hospital, proportion of patients under 15 years-old, proportion of patients above 60 years-old, proportion of female patients, students FTE, percentage of revenue spent on research, market forces factor.	Estimates of relative hospital efficiency are sensitive to estimation decisions and that little confidence can be placed in the point estimates for individual hospitals. The use of frontier techniques to set annual performance targets should be avoided.

FTE – Full-time equivalent; CMI – case-mix index; LOS – Length of stay.

Functional form of the production function

Regarding the structure of the production function, 14 studies used the Cobb-Douglas functional form, 16 considered the Translog and 6 studies estimated efficiency using both forms to test which one was a better fit. The remaining 7 studies used specific functional forms developed in previous works or developed a specific function for the study.

Apart from the studies of Rosko (2001a; 2001b) that reject the Cobb-Douglas functional form, inefficiency estimates from other studies are not affected by the choice of the functional form of the production function.

Error term distribution

What distinguishes SFA from DEA lies in the composition of the error term that besides inefficiencies also has a component that reflects random shocks out of hospitals control (Aigner, Lovell, and Schmidt 1977; Coelli, Prasada Rao, and Battese 1998; Li and Rosenman 2001; Jacobs, Smith, and Street 2006). The distribution of the error term should be defined a priori, and most of the studies opt to test which is the distribution that better fits the data.

Apart from some studies that have specific error distributions (Carey and Dor 2008; Farsi 2008; Kumar 2010) or time invariant error (Wagstaff and López 1996; Ludwig, Groot, and Van Merode 2009; Ludwig, Van Merode, and Groot 2010) all other studies test more standard distributions: half-normal, exponential and truncated. From these three the most tested distribution was the half-normal that was used in 25 studies. Exponential and truncated distributions were used in 11 and 10 studies respectively.

Variables

When estimating a cost-function, the most used measure of costs was the total costs of the hospitals. In fact, 33 of the 39 works that estimated a cost-function used the total costs as dependent variable whether those total costs were adjusted by case-mix. Other costs used included variable costs, operating costs and average costs per case.

Measuring hospital outputs is challenging and still debated in the literature. To measure health outcomes or patients' health improvements is not straightforward. For this reason, other measures have been used, such as the number of patients treated. This alternative is easily measured and quantifiable and seen as representing the production of the hospitals. However, each patient discharged uses a different combination of resources and is different from any other given the individual characteristics of each patient treated. This uniqueness of each discharge demands adjustments for complexity in order to account for it (Zuckerman, Hadley, and Iezzoni 1994; Vitaliano and Toren 1996; Wagstaff and López 1996; Chirikos 1998b; Rosko 1999). In accounting for the production of the hospitals inpatient discharges, outpatient visits or other absolute number of services provided have been frequently used. The average length of stay and the total number of inpatient days have also been considered.

Concerning hospitals' available resources, number of doctors, nurses, specialists, employees, etc, are frequently used as measures of labour resources. Bed-related variables, such as the number of beds, are often used as measures of capital. Wages and prices are also frequently used when measuring resources available.

When estimating the production function, to include hospital characteristics that may influence hospitals' efficiency might prove difficult. However, if the cost-function is estimated other characteristics can be included in the analysis. Those comprehend quality measures such as mortality rates, and occupancy rates; hospital's characteristics such as teaching status, ownership (private or public, profit or not-for-profit, etc), urban or regional locations, elderly and female ratios, degree of competition, etc.; and hospital service complexity measured by different services availability, high-technology availability, case-mix index, etc.

Main findings

Although almost every study presents an average efficiency level for the hospitals being analyzed, these estimates are not comparable between studies. Not only data in analysis is different but also there may be discrepancies on the selection, measurement and collection of the vectors of variables considered. Even in studies collecting data from different countries comparisons are not straightforward. Results on the interactions between variables and efficiency levels, however, are of main interest not only for research purposes but also to decision makers (Table 5).

Table 5: Main findings

Variables used	Evidence
Complexity and severity of patients	Mixed
Competition between providers	Mixed
Hospital size	Mixed
Hospital type	Mixed
Human resources	Mixed
Ownership	Mixed
Quality of care	Mixed
Specialization	Mixed
Teaching status	Inefficiency

Regarding the complexity and severity of patients treated, hospitals with higher case-mix indexes are the less efficient according to Li and Rosenman (2001). Nevertheless, Rosko and Mutter (2008) state that quality and severity controls do not have an impact on efficiency estimates. On the other side, Izón and Pardini (2017) suggest that higher hospital costs may be a consequence of the implementation of quality protocols affecting cost-efficiency.

Linna (1998) reports variations in quality of care as a possible source of inefficiency. Moreover, Deily and McKay (2006) found a positive relationship between hospital mortality and inefficiency. Ludwig, Van Merode, and Groot (2010), however, found that efficiency cannot be explained by lack of quality.

Teaching hospitals are identified as the hospitals with the most severe cases by Rosko and Chilingerian (1999) and the most costly hospitals by Rosko (1999) and Farsi and Filippini (2006). After adjusting for complexity, Farsi and Filippini (2008) found evidence that teaching hospitals are less cost-efficient that their counterparts.

General hospitals are considered more efficient than county hospitals by Gannon (2005). On the topic of hospital's networks, Carey (2003) and Rosko and Proenca (2005) admit that networks are associated to higher efficiency levels. Specialized hospitals are more efficient according to Kumar (2010) but more costly according to Farsi and Filippini (2008). Carey, Burgess, and Young (2008) find no evidence of higher levels of efficiency in specialized hospitals.

The size of the hospitals is a characteristic that is frequently assessed. Findings on this topic are ambiguous. Farsi and Filippini (2006) state that small basic care hospitals have the longest hospitalizations and Wang, Zhao, and Mahmood (2006) state that the very same hospitals have higher inefficiency scores. There is evidence that larger hospitals are usually more costly (Farsi and Filippini 2006) and less labour intensive (J. Wang, Zhao, and Mahmood 2006). Vitaliano and Toren (1996) find that hospitals with more beds are more efficient which is exactly the opposite of the result found by Li and Rosenman (2001) and B. B. Wang et al. (1999) that state that smaller hospitals with less beds have higher efficiency scores. Mateus, Joaquim, and Nunes (2015) observe that a better management of beds' occupancy might improve efficiency.

The influence of human resources on the productive process is not identical across countries (Mateus, Joaquim, and Nunes 2015) but according to Izón and Pardini (2017) there is a positive correlation between full-time equivalent employees per adjusted day and cost-inefficiency.

According to Herr (2008), Herr, Hendrik, and Boris (2011) and Barros, de Menezes, and Vieira (2013) private hospitals are more efficient than public hospitals, a result that is contested by Chirikos (1998). Deily, McKay, and Dorner (2000), McKay, Deily, and Dorner (2002), Rosko 2004) and Herr (2008) observe that not for profit hospitals have higher levels of efficiency but Rosko (1999; 2001a; 2001b), Li and Rosenman (2001) and McKay and Deily (2005) and Deily and McKay (2006) research suggest that for profit hospitals are more efficient. Farsi and Filippini (2006) and Farsi (2008) did not find a significant difference in efficiency between different ownership statuses.

The time since creation of health maintenance organizations is related to decreases in inefficiency scores which suggests that institutions tend to improve efficiency as time goes by (Sari 2003; Rosko 2004); Moreover, inefficiency declines with competition according to Rosko (1999), Rosko and Chilingerian (1999) and Rosko (2001b). Izón and Pardini (2017) results indicate that, despite counterintuitively, market competition increases inefficiencies.

Discussion and conclusions

The objective of this paper was to identify the main characteristics of the efficiency assessment using SFA: methodological framework, variables used and main findings. We were interested on how hospital resources' and characteristics affect efficiency and what are the main limitations of the SFA method.

Although SFA was developed to estimate the optimal production function, this method is, nowadays, more used for the estimation of the cost-function. The methodological framework of SFA is similar across studies: studies usually select one of the two common production functions available, the Cobb-Douglas or the Translog, to use the same distributions of the error term and to select the same variables. Findings on how the different resource variables and hospitals' characteristics affect efficiency, however, are not consensual.

Although datasets are different, and variables might not be comparable due to different processes of selection, collection and measurement of variables some limitations on the use of SFA in the healthcare sector may affect the accuracy of SFA estimates and explain the discrepancies found in the results. Moreover, as Ferrier (2014) states, there is little explanation on how the choice of functional form may affect efficiency scores as well as the distribution of the error term.

Newhouse (1994) and Folland and Hofler (2001) referred as limitation the incapacity of the estimated frontier to capture the heterogeneity of hospital outputs. Since each hospital discharge is unique some aspects such as patient characteristics and the variations in the inputs combination are not captured by efficiency estimations. Adjustments are useful but not exhaustive and other characteristics such as faster response times and personalized interaction with the patients are left behind and may possibly distort the estimations (Newhouse 1994). Moreover, Newhouse (1994) refers the case-mix variations not being as random as thought. Ferrier (2014) suggests that studies using SFA may use different granular levels of activity homogenizing the units of analysis to reduce heterogeneity issues.

The omission of explanatory variables is considered another limitation to the efficiency estimation (Dor 1994; Newhouse 1994; Folland and Hofler 2001). Costs and prices measurement are specially challenging in healthcare sector and often are omitted from the analysis. The estimation of a common production function has as limitation the non-inclusion of structural differences between groups of hospitals (Folland and Hofler 2001).

All these omissions lead to model misspecification and distorted efficiency estimates (Folland and Hofler 2001).

Another shortcoming in efficiency estimation lies in the assumption of inefficiencies where they do not exist. Inefficiencies may just be a result of a misspecified objective function, omission of relevant variables or constraints in the production process that should not be attributed to inefficiencies (Street 2003).

To overcome some limitations, Dor (1994) and Skinner (1994) suggest the use of a panel data to allow the estimation of a fixed effect for each health facility which would control the difference between the hospitals and the best practice.

The financial constraints being faced by most of the countries demand a further understanding on what contributes to good efficiency practices that minimize waste. SFA has many virtues however further research is deemed as necessary to improve the robustness of the estimates in the health care sector.

References

Aigner, Dennis, C.A.Knox Lovell, and Peter Schmidt. 1977. "Formulation And Estimation Of Stochastic Frontier Production Function Models." Journal of Econometrics 6 (1): 21–37. doi:10.1016/0304-4076(77)90052-5.

Barros, Carlos Pestana, António Gomes de Menezes, and José Cabral Vieira. 2013. "Measurement of Hospital Efficiency, Using a Latent Class Stochastic Frontier Model." Applied Economics 45 (1): 47–54. doi:10.1080/00036846.2011.579061.

Bernet, P. M., M D Rosko, and V. G. Valdmanis. 2008. "Hospital Efficiency and Debt." J Health Care Finance 34 (4): 66–88.

Carey, Kathleen. 2003. "Hospital Cost Efficiency and System Membership." Inquiry : A Journal of Medical Care Organization, Provision and Financing 40 (1): 25–38. doi:10.5034/inquiryjrnl_40.1.25.

Carey, Kathleen, James F. Burgess, and Gary J. Young. 2008. "Specialty and Full-Service Hospitals: A Comparative Cost Analysis." Health Services Research 43 (5 P2): 1869–87. doi:10.1111/j.1475-6773.2008.00881.x.

Carey, Kathleen, and Avi Dor. 2008. "Expense Preference Behavior and Management 'outsourcing': A Comparison of Adopters and Non-Adopters of Contract Management in U.S. Hospitals." Journal of Productivity Analysis 29 (1): 61–75. doi:10.1007/s11123-007-0065-3.

Chirikos, T N. 1998a. "Further Evidence That Hospital Production Is Inefficient." Inquiry 35 (4): 408–16. doi:0046-9580.

———. 1998b. "Identifying Efficiently and Economically Operated Hospitals: The Prospects and Pitfalls of Applying Frontier Regression Techniques." Journal of Health Politics, Policy and Law 23 (6): 879–904. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=9866092.

Chirikos, T N, and Alan M Sear. 2000. "Measuring Hospital Efficiency: A Comparison of Two Approaches." Health Services Research 34 (6): 1389–1408. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1089087/.

Coelli, T.J., D.S. Prasada Rao, and G.E. Battese. 1998. An Introduction to Efficiency and Productivity Analysis. Boston: Kluwer Academic Publishers.

Deily, Mary E., and Niccie L. McKay. 2006. "Cost Inefficiency and Mortality Rates in Florida Hospitals." Health Economics 15 (4): 419–31. doi:10.1002/hec.1078.

Deily, Mary E., Niccie L. McKay, and Fh Dorner. 2000. "Exit and Inefficiency: The Effects of Ownership Type." Journal of Human Resources 35 (4): 734–47. doi:10.2307/146370.

Dor, Avi. 1994. "Non-Minimum Cost Functions and the Stochastic Frontier: On Applications to Health Care Providers." Journal of Health Economics 13 (3): 329–34. doi:10.1016/0167-6296(94)90032-9.

Farrell, M. J. 1957. "The Measurement of Productive Efficiency." Journal of the Royal Statistical Society. Series A (General) 120 (3): pp. 253–90. doi:10.1016/S0377-2217(01)00022-4.

Farsi, M. 2008. "The Temporal Variation of Cost-Efficiency in Switzerland's Hospitals: An Application of Mixed Models." Journal of Productivity Analysis 30 (2): 155–68. doi:10.1007/s11123-008-0105-7.

Farsi, M, and M Filippini. 2006. "An Analysis of Efficiency and Productivity in Swiss Hospitals." Schweizerische Zeitschrift Fur Volkswirtschaft Und Statistik/Swiss Journal of Economics and Statistics 142 (1): 1–37.

Farsi, M, and Massimo Filippini. 2008. "Effects of Ownership, Subsidization and Teaching Activities on

Hospital Costs in Switzerland.” Health Economics 17 (3): 335–50. doi:10.1002/hec.1268.

Ferrier, Gary D. 2014. “Improving The ‘usefulness’ of Stochastic Frontier Analysis for Health Care.” Economics and Business Letters 3 (1): 27. doi:10.17811/ebl.3.1.2014.27-34.

Folland, S., and R. Hofler. 2001. “How Reliable Are Hospital Efficiency Estimates? Exploiting the Dual to Homothetic Production.” Health Economics 10 (8): 683–98. doi:10.1002/hec.600.

Gannon, Brenda. 2005. “Testing for Variation in Technical Efficiency of Hospitals in Ireland.” Economic and Social Review 36 (3): 273–94.

Herr, Annika. 2008. “Cost and Technical Efficiency of German Hospitals: Does Ownership Matter?” In Health Economics, 17:1057–71. doi:10.1002/hec.1388.

Herr, Annika, Schmitz Hendrik, and Augurzky Boris. 2011. “Profit Efficiency and Ownership of German Hospitals.” Health Economics 20 (6): 660–74. doi:10.1002/hec.1622.

Herwartz, Helmut, and Christoph Strumann. 2012. “On the Effect of Prospective Payment on Local Hospital Competition in Germany.” Health Care Management Science 15 (1): 48–62. doi:10.1007/s10729-011-9180-9.

Hollingsworth, B. 2003. “Non-Parametric and Parametric Applications Measuring Efficiency in Health Care.” Health Care Management Science. doi:10.1023/A:1026255523228.

———. 2008. “The Measurement of Efficiency and Productivity of Health Care Delivery.” Health Economics. doi:10.1002/hec.1391.

Hollingsworth, B, P J Dawson, and N Maniadakis. 1999. “Efficiency Measurement of Health Care: A Review of Non-Parametric Methods and Applications.” Health Care Management Science 2 (3): 161–72. doi:10.1023/A:1019087828488.

Izón, Germán M., and Chelsea A. Pardini. 2017. “Cost Inefficiency under Financial Strain: A Stochastic Frontier Analysis of Hospitals in Washington State through the Great Recession.” Health Care Management Science 20 (2): 232–45. doi:10.1007/s10729-015-9349-8.

Jacobs, R, P C Smith, and A Street. 2006. Measuring Efficiency in Health Care: Analytic Techniques and Health Policy. Measuring Efficiency in Health Care: Analytic Techniques and Health Policy. doi:10.1017/CBO9780511617492.

Kumar, Sameer. 2010. “Specialty Hospitals Emulating Focused Factories.” International Journal of Health Care Quality Assurance 23 (1): 94–109. doi:10.1108/09526861011010703.

Li, T, and R Rosenman. 2001. “Cost Inefficiency in Washington Hospitals: A Stochastic Frontier Approach Using Panel Data.” Health Care Management Science 4 (2): 73–81. http://www.ncbi.nlm.nih.gov/pubmed/11393744\nhttp://link.springer.com/content/pdf/10.1023/A:1011493209102.pdf.

Linna, M. 1998. “Measuring Hospital Cost Efficiency with Panel Data Models.” Health Economics 7 (5): 415–27. doi:10.1002/(SICI)1099-1050(199808)7:5<415::AID-HEC357>3.0.CO;2-9.

Ludwig, M, W Groot, and F Van Merode. 2009. “Hospital Efficiency and Transaction Costs: A Stochastic Frontier Approach.” Social Science & Medicine 69 (1): 61–67. doi:10.1016/j.socscimed.2009.04.026.

Ludwig, M, Frits Van Merode, and Wim Groot. 2010. “Principal Agent Relationships and the Efficiency of Hospitals.” European Journal of Health Economics 11 (3): 291–304. doi:10.1007/s10198-009-0176-z.

Mateus, C., I. Joaquim, and C. Nunes. 2015. “Measuring Hospital Efficiency--Comparing Four European Countries.” The European Journal of Public Health 25 (suppl 1): 52–58. doi:10.1093/eurpub/cku222.

McKay, Niccie L., and Mary E. Deily. 2005. "Comparing High- and Low-Performing Hospitals Using Risk-Adjusted Excess Mortality and Cost Inefficiency." Health Care Management Review 30 (4): 347–60. doi:10.1097/00004010-200510000-00009.

McKay, Niccie L., Mary E. Deily, and Fred H. Dorner. 2002. "Ownership and Changes in Hospital Inefficiency, 1986-1991." Inquiry 39 (4): 388–99. doi:10.5034/inquiryjrnl_39.4.388.

Meeusen, Wim, and Julien van Den Broeck. 1977. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." International Economic Review 18 (2): 435. doi:10.2307/2525757.

Mutter, Ryan L., Michael D. Rosko, and Herbert S. Wong. 2008. "Measuring Hospital Inefficiency: The Effects of Controlling for Quality and Patient Burden of Illness." Health Services Research 43 (6): 1992–2013. doi:10.1111/j.1475-6773.2008.00892.x.

Newhouse, Joseph P. 1994. "Frontier Estimation: How Useful a Tool for Health Economics?" Journal of Health Economics 13 (3): 317–22. doi:10.1016/0167-6296(94)90030-2.

Pratt, William R. 2010. "What Does Free Cash Flow Tell Us about Hospital Efficiency? A Stochastic Frontier Analysis of Cost Inefficiency in California Hospitals." Journal of Health Care Finance 37 (1): 35–44. http://www.ncbi.nlm.nih.gov/pubmed/20973372.

Rosko, M D. 1999. "Impact of Internal and External Environmental Pressures on Hospital Inefficiency." Health Care Management Science 2: 63–74. doi:10.1023/A:1019031610741.

———. 2001a. "Cost Efficiency of US Hospitals: A Stochastic Frontier Approach." Health Economics 10 (August 1999): 539–51. doi:10.1002/hec.607.

———. 2001b. "Impact of HMO Penetration and Other Environmental Factors on Hospital X-Inefficiency." Medical Care Research and Review 58 (4): 430–54. doi:10.1177/107755870105800404.

———. 2004. "Performance of U.S. Teaching Hospitals: A Panel Analysis of Cost Inefficiency." Health Care Management Science 7 (1): 7–16. doi:10.1023/B:HCMS.0000005393.24012.1c.

Rosko, M D, and Jon A. Chilingerian. 1999. "Estimating Hospital Inefficiency: Does Case Mix Matter?" Journal of Medical Systems 23 (1): 57–71. doi:10.1023/A:1020823612156.

Rosko, M D, and R L Mutter. 2008. "Stochastic Frontier Analysis of Hospital Inefficiency: A Review of Empirical Issues and an Assessment of Robustness." Medical Care Research & Review 65 (2): 131–66. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=18045984.

———. 2011. "What Have We Learned from the Application of Stochastic Frontier Analysis to U.S. Hospitals?." Medical Care Research & Review 68 (1 Suppl): 75S – 100S. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=prem&AN=20519428.

Rosko, M D, and Jose Proenca. 2005. "Impact of Network and System Use on Hospital X-Inefficiency." Health Care Management Review 30 (1): 69–79. doi:10.1097/00004010-200501000-00010.

Sari, Nazmi. 2003. "Efficiency Outcomes of Market Concentration and Managed Care." International Journal of Industrial Organization 21 (10): 1571–89. doi:10.1016/S0167-7187(03)00073-0.

Skinner, Jonathan. 1994. "What Do Stochastic Frontier Cost Functions Tell Us about Inefficiency?" Journal of Health Economics 13 (3): 323–28. doi:10.1016/0167-6296(94)90031-0.

Street, Andrew. 2003. "How Much Confidence Should We Place in Efficiency Estimates?" Health Economics 12 (11): 895–907. doi:10.1002/hec.773.

Vitaliano, D., and M. Toren. 1996. "HOSPITAL COST AND EFFICIENCY IN A REGIME OF STRINGENT REGULATION." Eastern Economic Journal 22.

Vitikainen, Kirsi, Miika Linna, and Andrew Street. 2010. "Substituting Inpatient for Outpatient Care: What Is the Impact on Hospital Costs and Efficiency?" European Journal of Health Economics 11 (4): 395–404. doi:10.1007/s10198-009-0211-0.

Wagstaff, Adam. 1989. "Estimating Efficiency in the Hospital Sector: A Comparison of Three Statistical Cost Frontier Models." Applied Economics 21 (5): 659. doi:10.1080/758524897.

Wagstaff, Adam, and Guillem López. 1996. "Hospital Costs in Catalonia: A Stochastic Frontier Analysis." Applied Economics Letters 3 (7): 471–74. doi:10.1080/758540809.

Wang, Bill Binglong, Yasar A. Ozcan, Thomas T.H. Wan, and Jeff Harrison. 1999. "Trends in Hospital Efficiency among Metropolitan Markets." Journal of Medical Systems 23 (2): 83–97. doi:10.1023/A:1020585118381.

Wang, J, Z Zhao, and A Mahmood. 2006. "Relative Efficiency, Scale Effect and Scope Effect of Public Hospitals: Evidence from Australia." IZA discussion paper no. 2520. http://ftp.iza.org/dp2520.pdf.

Yaisawarng, Suthathip, and James F. Burgess. 2006. "Performance-Based Budgeting in the Public Sector: An Illustration from the VA Health Care System." Health Economics 15 (3): 295–310. doi:10.1002/hec.1060.

Zuckerman, Stephen, Jack Hadley, and Lisa Iezzoni. 1994. "Measuring Hospital Efficiency with Frontier Cost Functions." Journal of Health Economics 13 (3): 255–80. doi:10.1016/0167-6296(94)90027-2.

Measuring hospital efficiency – comparing four European countries (Work 5)

Céu Mateus
Inês Joaquim
Carla Nunes

European Journal of Public Health:
Feb; 25 Suppl 1: 52-58

2015

DOI: 10.1093/eurpub/cku222

Best published paper by Researchers from
***Escola Nacional de Saúde Pública* Award, 2015**

European Journal of Public Health, Vol. 25, Supplement 1, 2015, 52–58

doi:10.1093/eurpub/cku222

Measuring hospital efficiency—comparing four European countries

Céu Mateus, Inês Joaquim, Carla Nunes

Escola Nacional de Saúde Pública, Universidade Nova de Lisboa, Lisbon, Portugal

Correspondence: Céu Mateus, Escola Nacional de Saúde Pública, Universidade Nova de Lisboa, Lisbon, Portugal, e-mail: ceum@ensp.unl.pt

Background: Performing international comparisons on efficiency usually has two main drawbacks: the lack of comparability of data from different countries and the appropriateness and adequacy of data selected for efficiency measurement. With inpatient discharges for four countries, some of the problems of data comparability usually found in international comparisons were mitigated. The objectives are to assess and compare hospital efficiency levels within and between countries, using stochastic frontier analysis with both cross-sectional and panel data. **Methods:** Data from English (2005–2008), Portuguese (2002–2009), Spanish (2003–2009) and Slovenian (2005–2009) hospital discharges and characteristics are used. Weighted hospital discharges were considered as outputs while the number of employees, physicians, nurses and beds were selected as inputs of the production function. Stochastic frontier analysis using both cross-sectional and panel data were performed, as well as ordinary least squares (OLS) analysis. The adequacy of the data was assessed with Kolmogorov–Smirnov and Breusch–Pagan/Cook–Weisberg tests. **Results:** Data available results were redundant to perform efficiency measurements using stochastic frontier analysis with cross-sectional data. The likelihood ratio test reveals that in cross-sectional data stochastic frontier analysis (SFA) is not statistically different from OLS in Portuguese data, while SFA and OLS estimates are statistically different for Spanish, Slovenian and English data. In the panel data, the inefficiency term is statistically different from 0 in the four countries in analysis, though for Portugal it is still close to 0. **Conclusions:** Panel data are preferred over cross-section analysis because results are more robust. For all countries except Slovenia, beds and employees are relevant inputs for the production process.

Introduction

Based on a data warehouse (DWH) developed during the European Collaboration for Health Optimization (ECHO) project, funded by the European Union FP7 programme, with information on hospital discharges, hospital characteristics and socio-economic population variables from four partner countries (England, Portugal, Slovenia and Spain), it was decided to perform an assessment of hospital efficiency. With this DWH some of the problems of data comparability are mitigated and new opportunities for efficiency assessments and healthcare system comparisons between the participating European countries are created.

Efficiency can be presented as the best use of resources in production,[1] the extent to which objectives are achieved in relation to the resources consumed,[2] the production of as many possible outputs given an available set of inputs,[3] among others. All these definitions imply the use of available resources in the best possible way, whether it is in the production of more output, better quality output or in the achievement of defined objectives. A technical efficiency-driven approach—maximization of output for a given level of inputs—is adopted.

Stochastic frontier analysis (SFA)[4] is one of the most popular methods for estimating optimal production functions needed in the estimation of efficiency. It presents strengths and limitations, as with any other method, and there is no consensus regarding the best one.[5] SFA is a parametric method that requires the specification of the production function and which only allows for a single output.[4] However, the error term is considered to include not only the inefficiency of the institution but also random fluctuations that are out of the control of the institution.[4,6,7]

The main objective of this article is to model efficiency levels in hospitals, with cross-sectional and panel data, using SFA and understand which are the most relevant resources impacting on efficiency estimates in each of the countries being analysed. International comparisons of efficiency are a particular challenge of this article and allow a relevant contribution in this area of research. The data available include English, Portuguese, Spanish and Slovenian patient discharges and hospital characteristics. Although data used to develop this study are restricted to four countries, the methodology presented here can be replicated using similar sets of data from other countries.

Background

The measurement of efficiency has developed over recent decades and there is now an extensive literature. The measurement of efficiency starts from the economic definition of the production function, the maximum amount of output obtainable considering a given bundle of inputs and available technology,[6] and the idea developed by Farrell[3] that inefficiency may be defined as the distance between a firm's output and the isoquant (the combination of different levels of inputs that produce the same level of output) of fully efficient firms or, in other words, the optimal production frontier.[4] The key point is that in practice the optimal production function is not known in practice and must therefore be estimated.

SFA was developed simultaneously by Aigner et al.,[6] and Meeusen and van Den Broeck,[8] and is a parametric methodology where the production function is estimated through a function such as Cobb–Douglas.[4] SFA assumes that differences between the actual performance of the institution and the optimal performance are not only the result of inefficiencies but also are due to random factors that healthcare institutions do not control. The error term is then defined as a sum of two components, a one-side distribution term that represents the specific inefficiency term and a normal distributed term that represents random fluctuations which are beyond hospital control.[4,6,7]

A comprehensive review of the literature on the estimation of hospital efficiency using SFA is provided in the work of Mateus et al.,[9] where 41 studies are analysed, namely on the *a priori* selection of the structure of the technology and the distribution of the error term, decisions that, according to tests performed by several authors, do not affect efficiency estimates.[10,11] The extension to a panel data model is invoked to introduce robustness into the analysis and should be used over cross-sectional analysis whenever possible.[7,12–15]

Most of the studies consider total costs as an output measure and only one study[16] considers cases weighted by the national average length of stay (as a proxy for the intensity of the resources used) as output. The input variables used to estimate efficiency are similar across studies and include measures of labour and capital (number of physicians, nurses, beds, etc.) and other characteristics such as quality measures, service complexity, teaching status, etc.

Frequently, findings on efficiency levels serve mainly as a reference since they are not comparable between studies. Nevertheless, in each study, evidence on which variables influence efficiency provides valuable information.

Generally, the main results are ambiguous. Case-mix, severity, quality, teaching status, size of hospitals, ownership, specialization and the time since the creation of the healthcare institutions were analysed by several authors. Positive or negative relationships with efficiency have been found by some authors while others have not found any relationship at all. A handful of limitations are, however, identified in the application of efficiency estimation methods in the healthcare sector.

The first limitation, referred by Newhouse[17] and Folland and Hofler,[18] is the inability of the estimated frontier to capture the heterogeneity of hospital outputs. The idea behind this limitation is that each hospital discharge is not equal to another not only because the patients' characteristics are different but also because they are a result of the combination of different inputs. Of course, outputs can be adjusted by complexity and severity, but this does not prevent other differences not captured by the model, such as faster response times or more interaction with patients, from distorting the findings.[17] Moreover, even when controlling for case-mix differences much of case-mix variations might not be at random.[17]

Another limitation of efficiency measurements is that there are usually potentially important inputs not included in the analysis.[12,17,18] This issue is especially relevant in healthcare efficiency analysis given the difficulties in input price measurements. Additionally, by estimating a common production function, structural differences between groups of hospitals are not taken into account.[18]

Altogether, these limitations may lead to model misspecifications and consequently distortions in the estimated efficiency.[18] Suggestions to improve the analysis include extending it to panel data since that allows the estimation of a fixed effect for each health facility, in order to control the difference between the hospitals and the best practice.[12,19]

Assuming inefficiencies where they do not exist is another common issue in efficiency measurement. As Street points out, inefficiencies detected might be a result of an incorrectly specified objective function, a failure to account for relevant inputs or a lack of recognition of the constraints on the production process.[20] More specifically, the error term usually interpreted as inefficiency might be the result of issues other than inefficiency. Transforming the variables into logarithms results in a normalization of the ordinary least squares (OLS) residual, and thus SFA becomes redundant since the normalization of the random error should be interpreted as random fluctuations beyond hospital control. On the other hand, heteroscedasticity might be the reason for the error term and even when models are corrected, inefficiencies might be attributed to heteroscedasticity and not be true inefficiencies.[20]

All the issues described earlier lead to the conclusion that the measurement of inefficiencies might not necessarily produce precise estimates.[13,19,21] This concern is especially relevant when there is an intention of translating inefficiency estimations into budgetary adjustments.[22] This limitation, however, does not hamper the usefulness of the methods as descriptive and analytical tools.[22] In fact, estimates may be useful for ranking hospitals, which is a more robust analysis.[21]

Data

ECHO's DWH population were public hospital discharges occurring in the consortium countries during the period under analysis. Thus, as for output, complete sets of inpatient discharges were available from 2002 to 2009 for Portugal, from 2005 to 2009 for Slovenia, from 2003 to 2009 for Spain and from 2005 to 2008 for England. The longitudinal availability of data enabled the extension of the analysis to a panel data approach. In all the countries, there is universal coverage of the population in health care and access to public hospitals is not dependant on people's wealth and funding is mainly obtained from public sources.[23–26]

Portugal and Spain used International Classification of Diseases-9th version-Clinical Modification (ICD-9-CM) to code both diagnoses and procedures and discharges were grouped into All Patient-Diagnosis Related Groups (AP-DRGs). Slovenia and England coded diagnoses with ICD-10. However, Slovenia used the Australian Classification for Health Interventions (ACHI) to code procedures, and discharges were grouped into Australian Refined-Diagnosis Related Groups (AR-DRGs). As for England, the OPCS Classification of Interventions and Procedures was used to code procedures and cases were grouped into Healthcare Resource Groups (HRGs). Groupers and coding systems do not compare directly. However, the purposes of the different groupers are similar: clinical coherence and identical consumption of resources for the cases in the same group. The resources considered are human and technical resources used in the provision of care to inpatient, and their combination is country dependant. The combination of those resources is taken into consideration to set the relative weights of DRGs or HRGs in each country. The discharges were weighted by the relative weights associated with the grouper used in each country, as an accurate way to adjust for the case-mix of each hospital in each country, this methodology being identical to the one selected by Gannon.[27] In England, instead of relative weights, a tariff expressed in pounds is associated with each HRG. To make similar adjustments to hospital discharges, tariffs were converted to relative weights. It was considered that the HRGs tariffs are composed of a relative weight, relating the average value of treating a patient within the HRG compared with the average value of treating all the patients, and a base rate, a value that when multiplied by the relative weights returns the tariff associated to each HRG. Assuming that the base rate is the average tariff paid to treat a patient, it was possible to get a relative weight to each HRG by dividing the tariff of the HRG by the base rate.

For the same period, four variables of hospital resources were considered as inputs in the analysis: the number of beds, as a proxy for capital inputs and headcounts of employees, physicians and nurses, as a proxy for labour inputs.

Over the period analysed, data correspond to 163 different hospital units in England, 102 in Portugal, 19 in Slovenia and 287 in Spain. Concerns might arise regarding the small number of decision making units available for the analysis in Slovenia and the results must be interpreted with caution. Nevertheless, Filippini et al.,[28] apply an identical methodology to five electricity distribution companies in Slovenia.

Methods

Following Aigner et al.,[6] the SFA model to estimate the production function using cross-sectional data is specified as follows:

$$y_i = f(x_i; \beta) + \varepsilon_i \text{ and } \varepsilon_i = u_i + v_i \quad (1)$$

where y_i is the output of the hospital (weighted hospital discharges), f is the functional form of the technology, x_i is the vector of inputs to be included in the analysis and β is a vector of parameters to be estimated.

The error term is composed of two terms: a symmetric disturbance v_i that is independently and identically distributed as $N(0,\sigma_v^2)$ represents the random fluctuations beyond hospital control; the inefficiencies of the hospitals are captured by the other term u_i that is independent of v_i and is independently and identically distributed as $N(\mu,\sigma_v^2)$.[4,6]

Technical efficiency is estimated as the ratio of observed output relative to the potential output of the hospital given its input level and represents the percentage of the production function attained by the hospital[4]:

$$\mathrm{TE}_i = \frac{y_i}{\exp(x_i\beta)} = \frac{\exp(x_i\beta - u_i)}{\exp(x_i\beta)} = \exp(-u_i) \tag{2}$$

In this study, the Cobb–Douglas functional form is chosen over the translog given the size of the sample ($n = 19$ in the Slovenian analysis, $n = 56$ in 2009 in the Portuguese analysis, $n = 161$ in 2005 in the English analysis and $n = 241$ in 2005 in the Spanish analysis).

As for the choice of the distributional form of the one-sided error term, given that there is no justification for the selection of any particular distribution,[4] three distributional forms—half-normal, truncated normal and exponential—were used in the estimation of the model.

In the panel data extension, these assumptions are relaxed and the performance of hospitals is also controlled for time period t, the inefficiencies of hospitals somehow being related with past inefficiencies. The production functions estimated using panel data can be specified as follows:

$$y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + v_{it} - s u_{it} \tag{3}$$

where y_{it} is the natural logarithm of the production of hospital i at time t, x_{it} is the vector of the natural logarithms of the inputs to be included in the analysis, β_j is the vector of the parameters to be estimated, v_{it} is the idiosyncratic error and u_{it} is a time-varying panel-level effect. If a time-invariant specification is selected then:

$$\begin{aligned} u_{it} &= u_i \\ u_i &\sim \mathbb{N}^+(\mu,\sigma_u^2) \\ v_i &\sim \mathbb{N}^+(0,\sigma_v^2) \end{aligned}$$

If a time-varying decay specification is selected then:

$$u_{it} = \exp\{-\eta(t - T_i)\} u_i$$

where T_i is the last period considered in the analysis and η is the decay parameter and:

$$\begin{aligned} u_i &\sim \mathbb{N}^+(\mu,\sigma_u^2) \\ v_i &\sim \mathbb{N}^+(0,\sigma_v^2) \end{aligned}$$

The decay parameter η gives information on the evolution of the inefficiency, i.e. if it increases or decreases over time. If η tends to 0, then the time-varying decay model reduces to a time-invariant model.

Given the limitations of efficiency measurements in health care, tests on the suitability of the available data are realized. In the cross-sectional case, two assessments are made[20]: (i) on the distribution of the OLS residuals and (ii) on the heteroscedasticity.

Following Street,[20] transforming the variables into logarithmic form normalizes their distribution, and consequently, the residuals are likely to approach a normal distribution. If the OLS error term follows a normal distribution, it implies that the residual variance should be interpreted as noise and not as inefficiency.[20] A Kolmogorov–Smirnov test for the normality of the distribution is performed for the variables included in the analysis. This test is also performed for the error term of the estimated OLS model.

The second assessment refers to the presence of heteroscedasticity because it can result in inefficiencies in the error term that are not real. The Breusch–Pagan/Cook–Weisberg test was used to test the presence of heteroscedasticity which may result in an underestimation of the intercept and an overestimation of the slope coefficients.[20]

The likelihood ratio (LR) test on the value of the one-sided non-random component of the error term (H0: $\sigma_u^2 = 0$) is also performed. This test can be performed both in the cross-sectional and panel data cases. If the null is not rejected, then σ_u^2 is not statistically different from 0 and the error term is considered to include only the random component of the error term and differences between the observation and the production function estimated are attributed to random shocks beyond hospital control. In this case, u_{it} can be removed from the model specification and SFA is not required because the parameters can be estimated using OLS.[13,29,30]

Given that only inpatient cases were considered in this analysis and that not all the remaining processes from a hospital—day hospital, ambulatory surgery, consultations, etc.—were included, the production functions obtained are only partial. For this reason, the number of employees, physicians and nurses were adjusted to reflect the hospital resources used in inpatient cases. The number of beds was not adjusted because it was assumed that all inpatient beds are exclusively allocated to inpatient services, which were the ones being considered in the current analysis.

Therefore only a percentage of human resources was allocated. Several values were tested for the percentage of employees, physicians and nurses: (i) 100% meaning that all the staff were only allocated to inpatient activities; (ii) hospital percentages that correspond to discharges over the total number of inpatient and ambulatory surgery discharges and (iii) 60% in order to account for labour time spent in outpatient visits, emergency visits, day hospital and other staff activities. The results with the different values can be found in the annexes and do not present significant differences. It was decided to use the percentages obtained for each hospital as described in (ii).

As pointed out, cross-sectional data might not be suitable to perform SFA. There is some evidence that the estimation of the productivity function using panel data should take into account the inefficiency term and that to specify a model with SFA is appropriate and might produce more reliable estimates. In order to accommodate different levels of efficiency and to be able to compare dispersion between countries, the z-score for efficiency levels in each country was computed.

Results

For Portugal, the reduction in the number of institutions from 2002 to 2009 reflects the reorganization of the Portuguese healthcare system which, since 2000, merged over 50% of the hospitals. The institutions still exist physically but are considered as one single organization where resources and production are shared. This contributed to the increase in the values of resources observed for Portugal. The high values of the standard deviations reflect the great heterogeneity in Portuguese hospitals—some small regional hospitals producing fewer than 1000 discharges per year and some hospital centres that aggregate several hospitals treating more than 50 000 patients per year.

In Slovenia, the number of hospitals is stable, and the average values for the variables (number of employees, physicians, nurses and beds) are also stable over the years of the analysis. Slovenian hospital reality is heterogeneous, as in Portugal, composed of some small hospitals that treat fewer than 1000 patients per year, while the largest hospital treats more than 100 000 adjusted hospital discharges each year.

Table 1 Descriptive statistics for the latest year*

Hospitals		Weighted hospital discharges			Employees	Physicians	Nurses	Beds
		Average (SD)	Min.	Max.	Average (SD)	Average (SD)	Average (SD)	Average (SD)
Portugal								
	56	18 831 (16 661)	941	69 553	1675 (1456)	297 (313)	564 (481)	404 (311)
Slovenia								
	19	24 661 (34 772)	755	151 766	692 (1055)	138 (247)	423 (610)	412 (522)
Spain								
	276	22 376 (22 444)	126	101 713	732 (711)	268 (251)	464 (468)	369 (344)
England								
	163	83 402 (47 096)	1805	230 538	2191 (1366)	514 (339)	1.029 (646)	727 (407)

*2009 is the latest available year in the case of Portugal, Slovenia and Spain; 2008 in the case of England.
SD, standard deviation.

Spain is the country with more hospitals in the analysis. Nevertheless, the average number of weighted hospital discharges is similar to the average weighted hospital discharges of Portugal and Slovenia (slightly more than Portuguese hospitals and slightly less than Slovenian hospitals). Once more, the standard deviation (and the minimum and maximum values of the weighted hospital discharges) reflects a great discrepancy in hospital size.

English hospitals present the highest average number of weighted hospital discharges. The standard deviation of weighted hospital discharges, contrary to the other countries, is lower than the average. As for resources, English hospitals present higher numbers of nurses and beds per hospital.

Table 1 summarizes the statistics of the variables included in the analysis. Results for all the years are presented in the Annexes. In the cross-sectional framework, both OLS and SFA models were estimated (Table 2) for all the countries. The results of SFA estimations using different resource adjustments can be found in the annexes. The existence of the non-random component of the error term in the SFA is tested using the LR test, also in Table 2. The Kolmogorov–Smirnov test was used to assess the normality of the unstandardized residuals of the OLS model and heterocedasticity is assessed using the Breusch–Pagan/Cook–Weisberg test.

For Portugal, SFA was estimated assuming half-normal distribution of the error since truncated and exponential distributions of the error term did not produce convergent results. For Slovenia, Spain and England, the results presented are the estimates when assuming the exponential error term distribution, since truncated and half-normal estimates did not converge. The results from the LR test reveal that, for Portuguese data, the results from the SFA and OLS models are not statistically different. This means that the one-sided component of the error term is not statistically different from zero and thus differences between the optimal production function and the activity of the hospitals can be attributed to random factors affecting the hospitals. In Slovenia, Spain and England, the results from OLS and SFA are, indeed, statistically different. Between 2002 and 2009 beds are always the most important variable for efficiency results for Portugal (Table 2), followed by nurses and physicians. The number of employees is never significant, but in 2002 the coefficient had a positive sign contrasting with the negative sign it showed in 2009.

For Slovenia, looking at the results of SFA, beds are also the most important variable to impact on efficiency measurements. The total number of employees is equally important in 2005. Physicians had a significant and negative sign in 2005, ending up still significant for efficiency analysis in 2009 but with a positive sign. The inverse can be observed for the nurses.

In Spain, the total number of employees had a coefficient with a negative sign in 2003 and became the variable with the highest coefficient in 2009 in the SFA. However, at the end of the period, both physicians and nurses presented a significant but negative coefficient. The number of beds was significant in every year of the period analysed and always positive.

The number of beds has a positive coefficient and is usually significant in the three countries presenting similar values. The number of physicians presents an identical trend for Portugal and Slovenia, being significant and with a positive sign in 2009. However, for Spain it shows a significant but negative sign in 2009. The number of nurses is always positive for Portugal; however, for Spain and Slovenia the coefficient can be positive or negative, depending on the years. For England, as well as for the other countries, the number of beds was significant in every year of the analysis contributing to efficiency measurement. Employees, physicians and nurses are significant in some years of analysis, although not consistently.

The number of beds has a positive coefficient and is usually significant in the four countries presenting similar values. The number of physicians presents an identical trend for Portugal and Slovenia, being significant and with a positive sign in 2009. However, for Spain it shows a significant but negative sign in 2009, being non-significant in the SFA analysis for England. The number of nurses is positive for Portugal and England; however, for Spain and Slovenia, the coefficient is significant but negative.

The results for Portugal, England and Spain are validated by the results of the Kolmogorov–Smirnov test. The test for Portugal estimates does not reject the null hypothesis of normality ($P > 0.05$, in all years), which suggests that performing SFA would be redundant since all variations might be attributed to randomness, as suggested previously by the LR test. The test on English and Spanish estimates does reject the hypothesis of normality ($P < 0.001$, in all years). This means that there is, in fact, evidence that the unstandardized residuals of the OLS estimates have more than just a normal distribution and that variation might not be attributed to randomness. For Slovenia, the result of the Kolmogorov–Smirnov test suggests that the unstandardized residuals follow a normal distribution but the LR test suggests that OLS and SFA are statistically different.

In order to test for heteroscedasticity, after the estimation of the OLS model, the Breusch–Pagan/Cook–Weisberg test was applied. While Portuguese data reveal homogeneous variance of the error term for all years except 2007 and 2009, both Slovenian, English and Spanish estimates present evidence of the presence of heteroscedasticity for all years ($P < 0.002$, in all years).

Once again the results show that efficiency measurements using frontier estimation when using cross-sectional data available might not be the best option, since those estimates might not be accurate. The Portuguese estimates using SFA with half-normal distribution

Table 2 OLS and SFA for the recent years

	England (2008)		Portugal (2009)		Slovenia (2009)		Spain (2009)	
	SFA[a]	OLS	SFA[b]	OLS	SFA[a]	OLS	SFA[a]	OLS
Constant	5.310***	4.826***	3.761***	3.759***	5.293***	7.865***	2.154***	1.387***
Beds	0.877***	0.850***	0.703***	0.703***	0.569***	−0.289	0.613***	0.557***
Employees	−0.093	−0.245	−0.344	−0.344	0.030***	−1.475	2.824***	3.018
Physicians	0.031	0.225*	0.274**	0.274**	0.352***	1.325**	−0.718**	−0.670**
Nurses	0.133	0.215 *	0.465**	0.465**	−0.053***	1.217	−1.651***	−1.770**
Log-likelihood	41.344	–	17.986	–	7.430	–	−53.427	–
LR test	38.84	–	0.000	–	17.580	–	120.000	–
***P*-value**	0.000	–	1.000	–	0.000	–	0.000	–

[a]Exponential distribution.
[b]Half-normal distribution.

Table 3 Results of the time-invariant and time-varying decay SFA models with panel data

	England						Portugal						Slovenia						Spain					
	TI			TVD			TI			TVD[(a)]			TI[(a)]			TVD[(a)]			TI[(a)]			TVD[(a)]		
N	649						574						95						1.814					
No. of groups	163						102						19						287					
	Coefficient			Coefficient			Coefficient			Coefficient			Coefficient			Coefficient			Coefficient			Coefficient		
Constant	5.349***			5.384***			3.776***			75.478			5.703***			5.296***			4.465***			4.413***		
Beds	0.094**			0.171***			0.488***			0.559***			0.021			0.042			0.436***			0.462***		
Employees[(b)]	0.542**			0.443**			0.312***			0.283***			0.808			1.141			0.648**			0.644**		
Physicians[(b)]	0.209**			0.189**			0.015			0.028			0.205			0.103			−0.642			−0.047		
Nurses[(b)]	0.074			0.115			0.264***			0.211***			−0.225			−0.459			−0.073			−0.104		
Log-likelihood	450.284			456.432			443.421			453.930			72.227			72.420			−90.875			−87.150		
Generalized R^{2}[(c)]	0.498			0.315			0.653			0.666			0.470			0.259			0.308			0.293		
		95% CI			95% CI			95% CI			95% CI			95% CI			95% CI			95% CI			95% CI	
η	-	-	-	0.037	0.016	0.058	-	-	-	0	0	0	-	-	-	−0.013	−0.053	0.028	-	-	-	0.016	0.005	0.026
σ^2	80.925	80.268	81.587	68.984	68.42	69.553	0.037	0.029	0.047	0.037	0.029	0.048	151.27	147.65	154.98	151.52	147.94	155.18	217.95	216.6	219.3	196.69	195.48	197.9
σ_u^2	80.918	80.258	81.577	68.977	68.411	69.544	0.03	0.021	0.039	0.031	0.022	0.04	151.26	147.6	154.93	151.51	147.89	155.13	217.91	216.56	219.26	196.64	195.43	197.85

TI, time-invariant model; TVD, time-varying decay model; CI, confidence intervals.
* *P*-value < 0.10; ** *P*-value < 0.05; *** *P*-value < 0.001.
[a]Convergence not achieved. Results after 100 iterations.
[b]Proportion of inpatient discharges over total inpatient and ambulatory surgery discharges.
[c]Generalization of the coefficient of determination (R^2) proposed by Cox and Snell[31] that reflects the improvement of the full model over the intercept model:
$R^2 = 1 - \left\{\frac{L(M_{\text{Intercept}})}{L(M_{\text{Full}})}\right\}^{2/N}$ where $L(M_{\text{Intercept}})$ is the likelihood of the model without explanatory variables, $L(M_{\text{Full}})$ is the likelihood of the model with the explanatory variables and *N* is the number of the observations in the analysis.

are not statistically different from OLS estimates and the variance might be attributed to randomness, while estimates using exponential distribution of the error term are not convergent. The Spanish, English and Slovenian data, while presenting SFA estimates that are statistically different from the OLS estimation, show evidence of the presence of heteroscedasticity which is a concern because results might not be reliable.

Table 3 presents the results from both the time-invariant and time-varying decay models for all the four countries. Log likelihood results and the generalized R^2 proposed by Cox and Snell[31] are also presented, as well as the estimated values for the parameters η and σ_u^2 and the respective boundaries of the 95% confidence intervals.

The models estimated were not convergent, except when using English data and the time-invariant model for Portuguese data. The results presented in this study are the results after 100 iterations of the model, which were set by the authors. The estimates of σ_u^2, the non-negative part of the error term that represents inefficiency, vary across countries.

For Portuguese data, the estimates of this parameter are close to 0, although the 95% confidence interval does not include it. This suggests that the differences between models, including an inefficiency term or not, are close to not being statistically significant. Nevertheless, the inefficiency term is statistically different from 0. For Slovenian, English and Spanish estimates, the parameter σ_u^2 is clearly different from 0 in both models and the inefficiency term should be included.

Regarding the time variation of the inefficiency term, for all the four countries the parameter η is close to 0, suggesting that the time-varying decay model can be reduced to the time-invariant model. In the four datasets, however, the log likelihood is higher in the time-varying decay models. The results of the generalized R^2, however, suggest that the level of improvement in the model by including the explanatory variables is higher in the time-invariant model in the Slovenian and Spanish analysis. Therefore, time-invariant models seem to explain the variability in weighted hospital discharges better than time-varying decay models. Portugal and England show opposite evidence.

Regarding the inputs of the production function, beds seem to play a significant part in the production process, except for Slovenia. The significance of human resources on the productive process varies across the four countries. Hospital employees, with the exception of Slovenia, are significant for both models. In England, physicians are significant in estimating the outputs, and nurses have statistical significance in the productive process in Portugal.

Efficiency levels are not directly comparable between countries. Nevertheless, within each country they can be assessed, as well as the dispersion of efficiency. Figure 1a and b present the dispersion of time-invariant and time-varying decay efficiency in each country. The evolution of efficiency in each country can be found in the annexes.

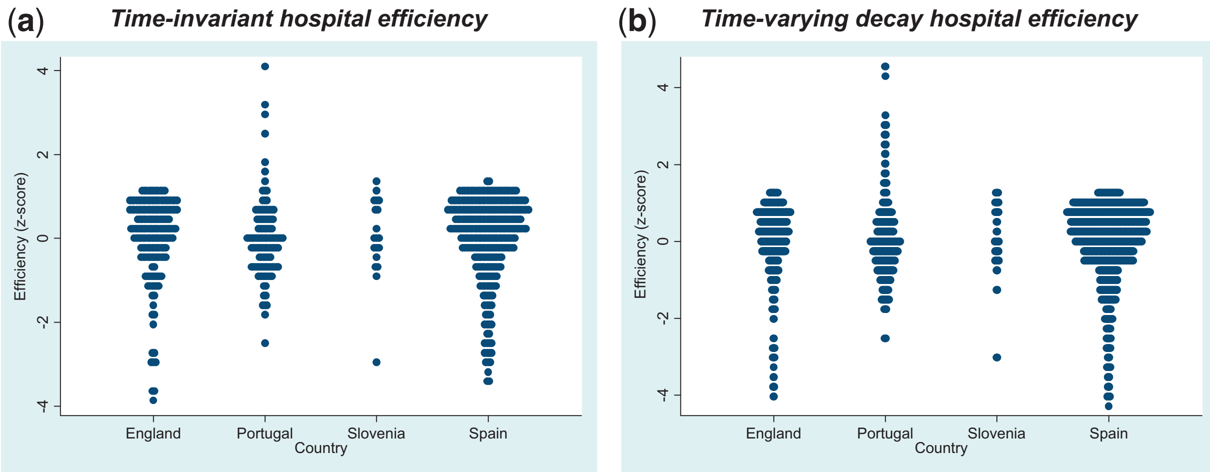


Figure 1 Turnip graphs for the (a) time-invariant and the (b) time-varying decay hospital efficiency (*z*-scores)

Efficiency dispersion is similar for Spain and England. Hospitals above the average are concentrated while hospitals below the average are more dispersed. This means that while there are not many outstanding hospitals, those above average have similar efficiency levels. There are also hospitals that have low efficiency levels when compared with other hospitals in their own country. In Portugal, differently from England and Spain, the difference between higher and lower efficiency levels is more pronounced. There are hospitals above the average, with higher levels of efficiency regarding the average, but also hospitals below the average level. However, a considerable number of hospitals present exceptionally high efficiency levels. The low number of hospitals in Slovenia does not allow for too much consideration but, as with Spain and England, outliers are more pronounced in low efficiency scores.

Discussion

The main objective of this article was to assess the suitability of the stochastic frontier analysis to estimate efficiency levels of hospitals using both cross-sectional and panel data and to appraise how beds and human resources impact on those levels. The number of hospital discharges adjusted for the complexity of the patients was selected as the output measure. Resources selected as inputs were the number of beds, employees, physicians and nurses. The limitations identified by several authors led to the performance of tests on the suitability of the data available to apply this methodology when using only cross-sectional data and panel data models.

Measuring hospital output is not straightforward given that the true final output—improvements in the health of patients—is difficult to assess in an objective manner in all the patients discharged. The number of patients discharged, however, is easily measured and quantifiable and can represent hospital production. Discharges can therefore be considered as a measure of hospital output although case-mix adjustments should be made.[11] In this study, the relative weights of each country were used to adjust respective inpatient discharges.

Recalling the results, Portugal presented normal distribution of the unstandardized residual, as well as SFA estimates that were not statistically different from the OLS estimates. As for the tests on heterocedasticity, the Breusch–Pagan/Cook–Weisberg test rejected the null only in 2007 and 2009. For Slovenia, although the SFA estimates were statistically different from the OLS estimates according to the LR test, the Kolmogorov–Smirnov test on the normality of the residuals was not rejected. Moreover, the Breusch–Pagan/Cook–Weisberg test found evidence of heterocedasticity, which suggests that SFA estimates are likely to be overestimated and not reliable. The results using Spanish data revealed that performing SFA or OLS models is statistically different. There is also evidence that the unstandardized residuals do not follow a normal distribution. However, there is evidence of heteroscedasticity, and thus the SFA estimates might not be accurate. Nevertheless, it is worth noting that the variable with the most similar coefficients in the four countries is the number of beds. The values obtained are positive and usually significant.

When extending the analysis to panel data, it was concluded that including the inefficiency term was indeed statistically different from not including it, and that SFA is a better option than a model specification that does not take inefficiency into account.

There is evidence that the productive process is significantly influenced by the availability of beds, except for Slovenia. This result was expected, and decision makers and hospital managers should take into account that a better management of beds may improve efficiency. Thus, measures that improve occupancy rates and help to reduce waiting lists will have a positive impact. The influence of human resources in the productive process is not identical in the four countries analysed. Hospital employees are significant in all countries but Slovenia. In England, physicians are statistically significant in the estimation of the optimal production function, while in Portugal nurses have statistical significance.

As for the dispersion of efficiency levels, a similar pattern seems to exist between Spain, England and, to some extent, Slovenia, where the outliers are the hospitals with exceptionally low efficiency levels. In Portugal, outliers are not only hospitals with extremely low efficiency levels but also outstanding hospitals. From the analysis performed one can say that there is room for improvement in efficiency levels in all countries.

With the results obtained in this study it has been shown that despite the difficulties in performing international comparisons on efficiency, it is possible to do so using data regularly collected and with simple methods such as OLS. When extending the analysis to panel data, the estimates of efficiency using SFA are reliable.

The results presented for Slovenia are, to our knowledge, the first assessing the efficiency of hospitals. Comparisons between efficiency levels in the different countries are not straightforward, but this study is a first attempt in the comparison of hospital efficiency between several countries. Panel data results are reliable for all the countries, and it has been shown that it is possible to evaluate efficiency levels using variables collected on a regular basis.

Supplementary data

Supplementary data are available at *EURPUB* online.

Funding

This work was funded by the FP-7 Programme of the European Union with Grant HEALTH-F3-2010-242189 for the ECHO project.

Conflicts of interest: None declared.

Key points

- International comparisons of the efficiency levels of hospitals are scarce due to problems related to the disparateness of data. However, we show that it is possible to evaluate efficiency levels using variables collected on a regular basis.
- The number of hospital discharges adjusted for the complexity of the patients was selected as the output measure. Resources selected as inputs were the number of beds, employees, physicians and nurses.
- In cross-sectional data SFA is not statistically different from OLS in Portuguese data, while SFA and OLS estimates are statistically different for England, Spain and Slovenia. Panel data should be preferred over cross-section analysis because results are more robust.
- The most relevant inputs for the production process are the number of beds and the number of employees. Decision makers and hospital managers should take into account that a better management of beds may improve efficiency.
- Measures that improve occupancy rates and help to reduce waiting lists will have a positive impact on efficiency levels.

References

1 Hollingsworth B, Dawson PJ, Maniadakis N. Efficiency measurement of health care: a review of non-parametric methods and applications. *Health Care Manag Sci* 1999;2:161–72.

2 Jacobs R, Smith P, Street A. *Measuring Efficiency in Health Care: Analytic Techniques and Health Policy*. Cambridge: Cambridge University Press, 2006: 243.

3 Farrell M. The measurement of productive efficiency. *J Royal Stat Soc (Series A)* 1957;120:253–90.

4 Coelli TJ, Prasada Rao DS, O'Donnell CJ, Battese GE. *An Introduction to Efficiency and Productivity Analysis*, 2nd edn. New York: Springer, 1999.

5 Lyroudi K, Glaveli N, Koulakiotis A, Angelidis D. The productive performance of public hospital clinics in Greece: a case study. *Health Serv Manage Res* 2006;19:67–72.

6 Aigner D, Lovell C, Schmidt P. Formulation and estimation of stochastic frontier production function models. *J Econometrics* 1977;6:21–37.

7 Li T, Rosenman R. Cost inefficiency in Washington hospitals: a stochastic frontier approach using panel data. *Health Care Manag Sci* 2001;4:73–81.

8 Meeusen W, van Den Broeck J. Efficiency estimation from Cobb-Douglas production functions with composed error. *Int Econ Rev* 1977;18:435–44.

9 Mateus C, Joaquim I, Nunes C. Estimating hospital efficiency using SFA-lessons from previous studies. Mimeo 2013.

10 Rosko MD, Mutter RL. Stochastic frontier analysis of hospital inefficiency: a review of empirical issues and an assessment of robustness. *Med Care Res Rev* 2008;65:131–66.

11 Zuckerman S, Hadley J, Iezzoni L. Measuring hospital efficiency with frontier cost functions. *J Health Econ* 1994;13:255–80.

12 Dor A. Non-minimum cost functions and the stochastic frontier: on applications to health care providers. *J Health Econ* 1994;13:329–34.

13 Rosko MD. Cost efficiency of US hospitals: a stochastic frontier approach. *Health Econ* 2001;10:539–51.

14 Linna M. Measuring hospital cost efficiency with panel data models. *Health Econ* 1998;7:415–27.

15 Sari N. Efficiency outcomes of market concentration and managed care. *Int J Industr Org* 2003;21:1571–89.

16 Herwartz H, Strumann C. On the effect of prospective payment on local hospital competition in Germany. *Health Care Manag Sci* 2012;15:48–62.

17 Newhouse JP. Frontier estimation: how useful a tool for health economics? *J Health Econ* 1994;13:317–22.

18 Folland ST, Hofler RA. How reliable are hospital efficiency estimates? Exploiting the dual to homothetic production. *Health Econ* 2001;10:683–98.

19 Skinner J. What do stochastic frontier cost functions tell us about inefficiency? *J Health Econ* 1994;13:323–8.

20 Street A. How much confidence should we place in efficiency estimates? *Health Econ* 2003;12:895–907.

21 Hadley J, Zuckerman S. The role of efficiency measurement in hospital rate setting. *J Health Econ* 1994;13:335–40.

22 Kooreman P. Nursing home care in The Netherlands: a nonparametric efficiency analysis. *J Health Econ* 1994;13:301–16.

23 Boyle S. United Kingdom (England): Health system review. *Health Syst Transit* 2011;13:1–483, xix–xx.

24 Barros P, Machado S, Simões J. Portugal: Health system review. *Health Syst Transit* 2011;13:1–156.

25 Albreht T, Turk E, Toth M, et al. Slovenia: Health system review. *Health Syst Transit* 2009;11:1–168.

26 García-Armesto S, Begoña Abadía-Taira M, Durán A, et al. Spain: Health system review. *Health Syst Transit* 2010;12:1–295, xix–xx.

27 Gannon B. Testing for variation in technical efficiency of hospitals in Ireland. *Econ Soc Rev* 2005;36:273–94.

28 Filippini M, Hrovatin N, Zorič J. Efficiency and regulation of the Slovenian electricity distribution companies. *Energy Policy* 2004;32:335–44.

29 Ludwig M, Groot W, Van Merode F. Hospital efficiency and transaction costs: a stochastic frontier approach. *Soc Sci Med* 2009;69:61–7.

30 Ludwig M, Van Merode F, Groot W. Principal agent relationships and the efficiency of hospitals. *Eur J Health Econ* 2010;11:291–304.

31 Cox DR, Snell EJ. *Analysis of Binary Data*, 2nd edn. London: CRC Press, 1989.

6. Discussion

Variations in medical practice exist in greater or lesser extent around the world. Their magnitude, their scope and their consequences are continuously being studied as a matter of interest for study of equity, quality of care and efficiency of healthcare systems and institutions. From the very beginning of the research on this topic, to the creation of international databases to enable international comparisons this research topic has grown in different ways.

This thesis had five specific objectives divided in three main sections – identification, causes and consequences – aiming at contributing to the topic of variations in medical practice. The first two objectives focused on the *identification* of variations in medical practice in Portugal given the lack of knowledge of this problem. The first objective was to identify the geographic variations in health care utilization in Portugal based on patients' place of residence for several procedures and the second objective was to focus on a specific procedure – the C-section – to quantify the geographical variability and excess use comparing different European countries.

The first objective was answered in the first work (W1) where a set of procedures based on high-cost, high-volume, policy relevance and data availability was selected and analyzed based on patients' place of residence. The procedures selected were: CABG, percutaneous transluminal coronary angioplasty (PTCA), catheterization, surgery after hip fracture, knee replacement, knee arthroscopy, C-sections per 1,000 live births and hysterectomies.

Results show that some procedures have seen their geographical variations reduced (such as catheterization, PTCA and C-sections to some extent) but for others (knee arthroscopy and more clearly knee replacement) geographic variations are persistently high. These results are consistent with the results of other European countries (Carinci, Stanislao e Moirano, 2014; Farebrother, 2014; Or e Verboux, 2014; Pellegrini e Kohler, 2014; Srivastava *et al.*, 2014).

It is important to note that geographical variations indicators should not be analyzed exclusively. The tendency of procedure rates (whether procedure activity is increasing or reducing), for example, supports the analysis by giving a signal of the robustness of the trend of the coefficient of variation. A reduction in the geographical variation combined with an increase in the volume of the procedure is a clear sign that geographical variations are, in fact, reducing and that it is expected that adoption and access to good clinical practice is occurring. This is the case of PTCA where the increase of procedure rates combined with the decrease of geographical variations across the countries suggests that good clinical practice is being implemented. The same tendency has been found in Canada (Alari, Lafortune e Srivastava, 2014).

Some caution should be placed on the accuracy and consistency of zip-code and area of residence of patients. Not always the zip-code is up to date and the treatment is not required to be performed in the hospital of the catchment area. Patients may choose where they wish to be treated. Analysis of zip-code compared with hospital of treatment reflect that 77% of the patients are being treated on their reference hospital. Of the remaining 23% of the patients, 61% are being treated in a tertiary hospital in metropolitan areas such as Lisbon and Oporto which may be due to specialization of treatments in these hospitals or are dislocated workers or students. Only 9% of the patients are being treated in hospitals out of their zip-code area.

The second objective – to quantify the geographical variability and excess use of C-sections comparing different European countries – is answered in the second work (W2) which compares the use of C-sections in Denmark, England, Slovenia and Spain. Two different indicators of C-section use were studied: standardized C-section rates (considering differences in age and sex population distributions) and lower-value indication C-section rates (considering only low-risk pregnancies and deliveries).

Total C-section rates have slightly increased across countries over the period of analysis. Lower-value indication C-section rates have been stable though with a downward trend in England and Spain and an upward trend in Denmark. Comparing countries, Portugal is the country with the highest rate of C-sections (32%) and Slovenia presents the lowest rate (19%). These results do not hold when analyzing lower-value C-sections where Denmark is the country with the highest rate (10%) and Portugal presents the lowest rate (2%).

Regarding geographical variations, Denmark is the most homogenous country followed by England and Slovenia. Portugal and Spain present high levels of geographical variations whether talking about standardized C-section rates or lower-value C-section rates.

Although Portugal presents the lowest rates in lower-value C-sections, it is the country, together with Spain which presents the highest excess consumption, around 80%. This indicator is estimated by computing the C-sections that could be avoided if all geographical areas realigned the activity by the lowest intensity area. Countries with highest variations have great percentages of care that could be avoided thus. Although Denmark is the country with lowest variation, it is estimated that there is between 30 to 37% of excess consumption of lower-value C-sections.

The results of W2 highlight three points that should be addressed: 1) the importance of being specific in defining the type of care; 2) the importance of international benchmarking.

It has been seen from W1 and W2 that two pairs of indicators related to C-sections provided different results between them. Defining the scope of the procedure is relevant when analyzing variations and excess consumption. In the specific case of C-sections, it is not the total C-sections that should be tackled but the C-sections that are being performed in lower-risk pregnancies. These are the ones that need to be minimized and to which the measures and policies to reduce them must work.

International benchmarking is starting, although conclusions are always conservative, and estimates remain “nationally based”. In W2, excess consumption in Denmark in lower-value C-sections is around 30% but rates on these procedures are much higher than in Portugal where excess consumption is estimated to be 80%. Would the excess consumption hold in Denmark would this country be compared to Portugal?

A drawback is the limitation of this study to NHS scope. However, for all the countries being studied this represents most of the deliveries (more than 95% in England and Denmark, around 90% in Portugal and Slovenia and around 80% in Spain) which does not suggest a bias on the results.

Each country is collecting hospital activity data but if data collection purposes and incentives are different across countries so can be coding practices. The information bias generated by different coding practices may jeopardize the comparisons between countries and therefore international comparisons remain conservative. Nevertheless, these initial international comparisons provide a glimpse on the different practices across countries.

The persistent geographic variations unravel the need to address and ensure that appropriate clinical assessment of the potential benefits and risks of a surgery are considered and ultimately that clinical guidelines are required (Birkmeyer *et al.*, 2013; Folland *et al.*, 2013). Especially when literature systematically indicates that the drivers of intensity of utilization of healthcare lies on the supply-side (Baicker, Buckles e Chandra, 2006; Lauer *et al.*, 2010).

Several strategies have been though to tackle geographical variations and promote good clinical practice. These include physicians’ training and information programs; monitorization and information on regional performances; targets setting at the regional levels to promote appropriate use of healthcare services; re-allocation of resources to correct for over or under treatment that is thought to be influenced by supply; implementation of clinical guidelines that homogenizes clinical practice; provider-level reporting and feedback; changes in payment systems to correct of over or under treatment; measurement of health outcomes; utilization of decision aids for patients promoting informed decisions and responsiveness to patients’ preferences (Folland *et al.*, 2013; Srivastava *et al.*, 2014; Wennberg e Fowler, 1977; Wennberg, 1984).

Tonsillectomy’s rates have been significantly affected by an informational program that was established in England (Folland *et al.*, 2013; Wennberg e Fowler, 1977). Pelvimetries performed without need were also reduced after an information program was established (Chassin e McCue, 1986; Folland *et al.*, 2013). In Canada, a review program reduced unjustified hysterectomies (Dyck *et al.*, 1977; Folland *et al.*, 2013).

In Spain, the result of an analysis on the geographical variations in C-section rates and the findings that variability could not be attributed to differences in need prompted the development of monitorization of a set of indicators of misuse, the development of clinical guidelines on normal delivery. As a result, C-sections rates have reduced and stabilized (Alvarez-Bartolomé e Gogorcena-Aoiz, 2014). The same happened in the United Kingdom, where the clinical guidelines provided by the National Institute for Health and Care Excellence (NICE) have kept C-section rates at an average level below OECD level with low geographical variability (Farebrother, 2014).

In knee arthroscopy, one of the procedures that is consistently in a high level of geographical variability across countries, there are some PROMs being developed to determine the benefits of these interventions to patients, whether they are high or only marginal (Farebrother, 2014).

It is generally concluded in the literature, that the identification of geographical variations in medical practice is the first step to address the topic. The studies identifying geographical variations show the type of information and metrics that can be applied to compare the utilization of procedures. This data is helpful for decision-makers who can screen the healthcare systems to identify and quantify excess-consumption on the utilization of specific procedures, to identify opportunity costs and to estimate local opportunities for improvement comparing geographical areas. Although excess use is usually the concern of policy makers, these studies can also provide information on geographical areas that are below the activity rates considered minimum thresholds for good quality of care.

The identification of geographical variations promotes the study of the causes of these variabilities as a way to reduce this phenomena (Or e Verboux, 2014; Schang *et al.*, 2014). When incentives are sorted out, effective policies to minimize and tackle lower-value care are easier to design. This is why understanding the cause behind variations in medical practices is so important.

The third objective of this PhD thesis was related with understanding the causes for variations in medical practice by studying whether resource availability could explain the excess use of C-sections in Portuguese NHS hospitals and to estimate the costs of the excess use.

W3 answers this objective by looking at the cesarean sections that were performed without an extended list of medical reasons and are thus considered “avoidable”. Analysis includes NHS deliveries in Portugal and is extended to understand how the rate of avoidable C-sections is influenced by the availability of medical resources and the economic burden of the excess consumption of this procedure.

C-section rates (CSR) have been steadily increasing between 2002 and 2009 from 26% to 31% with a turnover in 2010 reaching 29% in 2011 values that are way above the WHO threshold of 15%. Focusing on avoidable C-section rates (A-CSR), these rates have steadily increased between 2002 and 2008, starting to decrease in 2009 and continuing this decreasing trend until 2011.

Variation is studied at hospital level and there is a significant dispersion on the A-CSR (CV around 38%). It is assumed that the best hospital performance possible is 0% - in cases where CS would not be performed without any justification. The lowest A-CSR presented by a hospital is 5% and the highest is 40%. In total, in Portugal, avoidable C-sections (A-CS) are estimated to have an economic burden of €3.2M of wasted resources that could have been potentially saved or used to treat other patients if vaginal delivery had been chosen instead. This value represents 4% of the total cost of deliveries. Additionally, since the best care is not being delivered questions of the quality of care should also be considered.

Variability in A-CSR have been studied in the literature and justified by differences in medical practices, behaviors, preferences, hospitals characteristics and availability of resources. It is usually thought that the more the resources available the more the activity is done. Literature has demonstrated that it is not always true and that the type of resources available influences the results. Availability of resources may induce activity, while activity is limited when fewer resources are available (Aelvoet *et al.*, 2008; Brown, 1996). However, when specialized resources are considered, a wiser utilization is reported (Clark *et al.*, 1998).

From the four variables of hospital resources that were studied to understand their impact on the variability of A-CSR between hospitals, number of beds of hospital and teaching status are negatively correlated to A-CSR, availability of neonatal intensive care unit (ICU) is positively correlated and ratio of obstetricians to obstetric beds was not statistically significant in explaining A-CSR variability. Time trend is positive and statistically significant proving that regardless of hospital characteristics, A-CSR has increased over the years.

Interpretation of these results suggest that bigger hospitals are promoting wiser utilization of resources which may be a result of working at fuller capacity or that the most specialized resources are found in these hospitals. The positive correlation between availability of neonatal ICU care and the A-CSR suggests that availability of resources is increasing the A-CSR more than the potential specialization of resources is pushing that rate down. It is worth noting that hospitals with a neonatal ICU are also the bigger and more specialized hospitals.

Differences between CSR of teaching and non-teaching hospitals are not found in the literature (Aelvoet *et al.*, 2008; Snyder *et al.*, 2011). Despite the lack of evidence, expectations were that teaching hospitals would be better at following guidelines and best practices, thus presenting lower levels of A-CSR which was in fact what happened in the analysis with teaching status being negatively related to A-CSR. Nevertheless, 98% of the hospitals in the dataset are teaching hospitals which may be influencing the results.

These results are not truly comparable to the results in the literature mainly due to the different scope of analysis. Using A-CSR instead of using the general CSR or the Nulliparous, Term, Singleton, Vertex (NTSV) CSR makes the results not truly comparable. Incomparability also happens across countries due to the international discrepancies in the classification of deliveries without complications (Or *et al.*, 2012). Different models have been used in the literature with some authors modeling the probability of C-section delivery and other authors modeling hospital variations using fixed or random-effects models.

Despite the statistical significance of the tested resource variables, the little explanatory power of the estimated models suggests that much of the variations in A-CSR remain to be explained. This may mean that other variables of availability of resources (such as anesthetists) should have been included or that a significant part of variations in medical practice are not influenced by resource availability but by other factors such as clinical practice, organizational arrangements, mothers' preferences and economic incentives, data that was not available for analysis. This supports the idea that policies on resource reduction to improve efficiency may not be the best option. Instead, policy makers should focus on individual and institutions incentives and design policies that influence activity towards efficient practices.

The results found in the three previous works focused on the identification of variations in medical practice and in understanding the causes of these variations. After identifying the variations in medical practice and studying the causes of these variations, estimating the consequences of these variations will provide insight on the magnitude of the inefficiency that is generated by this phenomenon.

Methods for efficiency/inefficiency estimation include the DEA and the SFA. The SFA is a much more demanding method of efficiency estimation than the DEA is. However, SFA admits an error term for fluctuations out of hospital control which make it desirable when evolution of efficiency is being analyzed. The forth objective of this PhD thesis was to study the SFA as a methodology for efficiency measurement and its applicability in health setting, which variables are used, the main results and the limitations. The fifth objective was to apply this methodology to assess and compare hospital efficiency levels within and between four European countries.

In W4, a critical review of studies using SFA in healthcare setting is performed. Forty-one studies were identified and analyzed in terms of number of institutions considered, time period, structure of the production function, distribution of the error term, variables selected to be included in the analysis and main results.

The first step into SFA is the choice of the production function that emulates the healthcare production function. Cobb-Douglas and Translog production functions are frequently selected by researchers who also tend to select similar distributions of the error term – half-normal, exponential or truncated – as well as the same inputs – number of employees (doctors, nurses and/or specialists) or in alternative wages – and outputs – number of patients treated adjusted for complexity, the average length of stay or total number of patient days.

Since the healthcare sector is characterized by having diverse and heterogenous outputs researchers in this area tend to estimate the cost-function instead of computing the production function as it allows the inclusion of the different outputs produced combining them in monetary terms (Rosko e Mutter, 2011). When costs are not available, production function is possible to estimate, although it admits one single output only. The choice of a single output, though, is complex researchers may combine different outputs into an index or use relative weights to address differences in procedures' complexity. In our review, three out of forty-one studies estimated the production function. Gannon (2005) decided to aggregate inpatient discharges, outpatient visits and day cases in one single output, while Herr (2008) and Herwartz and Strumann (2012) weighted the cases to obtain comparable number of cases.

Most of the studies in this area concluded that the choice of the functional form of the production function did not affect the inefficiency estimates. As for the distribution of the error term, the majority of the authors tested which distribution better fitted the data with a special focus on half-normal, exponential and truncated distributions though some authors decide to use specific error distributions (Carey e Dor, 2008; Farsi, 2008; Kumar, 2010) or time invariant error (Ludwig, Groot e Merode, Van, 2009; Ludwig, Merode, Van e Groot, 2010; Wagstaff e López, 1996).

Research on this topic is not conclusive on how the different inputs such as complexity and severity of patients, quality of care, size of hospital, ownership, and specialization affect efficiency with authors having contradictory results.

Identified limitations to SFA include: i) the complexity to capture the heterogeneity of outputs into the estimated frontier (Folland e Hofler, 2001; Newhouse, 1994); ii) the omission of variables that are challenging to measure (Dor, 1994; Folland e Hofler, 2001; Newhouse, 1994). For these reasons, several authors state that conclusions on the efficiency estimates should be taken carefully (Dor, 1994; Hadley e Zuckerman, 1994; Skinner, Jonathan, 1994). Nevertheless, as stated by Kooreman (1994), SFA can be a useful as descriptive and analytical tool despite the limitations.

The W5 applies SFA methods to compare hospitals' efficiency levels in 4 European countries – England, Portugal, Slovenia and Spain – an objective that has been challenging to address and that was possible given the construction of ECHO international database. A panel data extension is performed to control hospital performances for time-period *t*, i.e., the inefficiencies of hospitals somehow being related with past inefficiencies.

Production functions for each country were estimated using a Cobb-Douglas functional form. The production of weighted hospital discharges (output) considered as inputs the number of beds, physicians, nurses and employees adjusted for the percentage of inpatient activity out of hospital total activity. The three more common error distributions (half-normal, exponential and truncated) were tested to assess which one would fit better. This resulted in using half-normal distribution for Portugal and exponential distribution for England, Slovenia and Spain since using other distributions did not produced convergent results.

Production function estimates and their respective errors' distribution suggest that Portuguese hospitals inefficiencies are attributed to random factors. Slovenian, Spanish and English hospitals present statistically significant positive error term suggesting that inefficiencies are present.

After Breusch–Pagan/Cook–Weisberg test was performed, results suggest some form of heteroscedasticity. This means that cross-sectional data available might not be the best option to perform SFA and that the estimates produced might not be accurate. Results from panel data estimates for time-varying decay and time-invariant models suggest that time-varying decay estimates are better for Portuguese and English data and time-invariant estimates are better for Slovenian and Spanish hospitals.

The number of beds is the most important resource to affect efficiency in Portugal affecting also efficiency in Spanish and in English hospitals. In panel analysis, this variable loses importance in the estimation of efficiency in Slovenia, though. As for the number of nurses, physicians and employees, these variables affect efficiency differently in the four countries. The importance of the number of beds in the production function is an indicator for decision makers and hospital managers who should consider that a better management of beds may improve efficiency. Measures that improve occupancy rates and help to reduce waiting lists may have a positive impact on efficiency levels.

Efficiency levels may not be directly comparable between countries, but some considerations can be done in the dispersion of efficiency results within countries. In Spain, Slovenia and England, hospital efficiency is concentrated above average performance without many outstanding hospitals. In Portugal, dispersion is higher with outstanding hospitals and low-performing ones. These results suggest that there is room for improvement in efficiency levels in all countries especially when looking at low performing hospitals.

Despite the several limitations identified previously in the literature and the limitations in applying SFA to ECHO database when extending SFA analysis to panel data, it was concluded that including the inefficiency term was indeed statistically different from not including it, and that SFA is a better option than a model specification that does not take inefficiency into account.

The results have shown that international comparisons are possible and that SFA produce reliable estimates when panel data is used. Nevertheless, comparisons between efficiency levels in these countries should be taken with caution.

Some additional limitations pertaining international database construction should be addressed since they may limit the analysis and the results.

Countries populating the ECHO database use different coding systems to register hospital activity. While diagnoses were registered in two different coding systems which are nevertheless related (ICD-9-CM and ICD-10), procedure codes registration used 4 different coding systems for 5 countries (ICD-9-CM, NOMESCO, OPCS and ACHI). Although crosswalks between coding system were performed this ends up being a limitation when selecting procedures to study because there are different coding practices. Apart from coding practices, it should also be mentioned the coding intensity differences between countries.

Another limitation is the heterogeneity of countries being studied in size and geographical units. Variations in medical practice are compared between these countries but some of the differences between variation levels may be related to population size differences and are not captured in the results.

Data collected usually refer to activity that is performed within the health system. This data lacks information about clinical outcomes and quality of life that result from the activity performed. This lack of information limits analysis on how variations in medical practice reflect on quality and efficacy of care provided by national health systems.

Going forward

Having in mind the key findings on the works produced, several questions arise that propel us forward in the analysis of variations in medical practice, efficiency estimation and overall analysis on the performance of healthcare systems.

The main output of the analysis of SAV and the identification of regional differences within and between countries is the confirmation of the disparities not only in healthcare access but also in quality of care. The methods used identify regional disparities in healthcare provisions and even the regions which are performing better or worse. With a defined optimal threshold, it is also possible to estimate the excess consumption and thus the current waste on specific procedures.

Considerations on the appropriate population to benefit from healthcare should be taken thoughtfully when addressing variations in medical practice, optimal threshold and excess consumption. Results may say little about the adequacy of healthcare provided if cohorts are not specifically defined or if population are broad enough to generate misinterpretations. Take the 15% optimal C-section rate threshold indicated by the World Health Organization (WHO). In extreme, a country may reach this optimal threshold using the procedure in women who do not need it while women in need are not provided the adequate care. In this extreme case, researchers could be stating that the country is reaching an optimal procedure rate while, in fact, healthcare provided was not the most adequate.

Different causes for variations in medical practice have been described in the literature and include medical practices, behaviors, preferences, hospitals characteristics and availability of resources. The data required to analyze the causes of variations in medical practice is not always available, though, resulting in modest explanations on the variations in medical practice.

When estimating the consequences of variations in medical practice, namely the magnitude of inefficiency, works focus essentially on the relationship between costs and outputs. Efficiency is usually measured focusing on the activity produced by healthcare institutions such as inpatient stay, length of stay or at the most analyzing quality of care. Often, this activity is translated into cost analysis given the flexibility and homogenization of activity that cost measures allow.

We, researchers, should be posing ourselves several questions:

- Are we using the correct measures to address patient access?
- Are age and sex standardization methods still sufficient to address differences in populations or should we be looking to more specific conditions of populations?
- Are activity measures, or cost measures sufficient to address good care or are we perpetuating the idea that more activity with fewer resources is better?
- Should we continue to consider more activity at a lower cost, better healthcare? What about the true results of the healthcare provided?
- Should we be chasing different data than the data that is being collected and that is currently available?

The methodologies used to identify variations in healthcare are indeed a starting point to perceive heterogeneity in healthcare provision. Conclusions on the heterogeneity are conditional on the scope of the analysis, though, with a deeper analysis on specific populations being required to provide good insight. A clear definition on the population benefiting from specific care is required and conclusions on patient access should consider the defined population.

As for the analysis of the quality and efficiency of healthcare systems, the relationship between the quantity of care produced with the resources available is still the focus when addressing these topics. Different indicators of good care should be included in the analysis. They are seldom available, though.

The works produced are the first step to understand problems of patient access, of efficiency and waste, but they are not the end.

Future research

A new paradigm is emerging after works of Michael Porter and other researchers have proposed a shift in the way society looks to healthcare systems (Porter e Lee, 2013; Porter e Teisberg, 2004).

In their words, healthcare systems have been struggling with the wrong broad goals, thus walking in the wrong direction (Porter e Lee, 2013). Until recently, increasing activity at the lowest possible cost or achieving a specific threshold of activity were the objectives set for different procedures. Increasing access to bad care, reducing costs at the expense of quality or providing an optimal threshold to the wrong population are not the objective, though. What healthcare system should be looking for is value of healthcare provided to patients and to deliver care that provide good results (Porter e Lee, 2013).

Value is defined as the outcomes that matter to patients relative to the cost of achieving these outcomes (Porter e Lee, 2013). The shift from output to outcome is an important one. It softens the importance of a specific procedure to make the good clinical result the main objective. The concept also combines in one single *number* results and costs. The idea of a ratio that puts together results and costs is not new. Cost-effectiveness analysis had already combined both. Difference is that cost-effectiveness analysis was used to compare two alternatives regarding results (benefits) and costs. It was used mainly in health technology assessments to compare alternative therapies or drugs. Value definition combines both outcomes and costs in a single measure not for comparisons but as a final purpose.

Variations in medical practice exist and furthermore, they exist combined with variations in medical costs and certainly with variations in medical outcomes (Erskine *et al.*, 2016).

The shift towards value changes the way we may look to the findings of this PhD thesis. Variations in medical outcomes will certainly provide more meaningful information on how we should be treating patients than knowing how much we should be treating. Variations in medical practice are the starting point for the study of value (Moriates, Arora e Shah, 2015).

One thing we have learned from literature is that increased healthcare spending or increased activity is not truly related to better results (Moriates *et al.*, 2015). Several authors have suggested different strategies to reduce variations in medical practice. One of them was related to homogenization of clinical practice through implementation of clinical guidelines.

Clinical guidelines are usually implemented according to the best literature and often after realization of cost-effective analyzes which would add economic analyzes to the clinical benefits of therapies. These guidelines would homogenize medical practice towards the most efficient care. Clinical guidelines, however, have been often implemented as recommendations for drugs while a new perspective on value of procedures, diagnoses and disease management is required. It is nevertheless required to create meaningful clinical guidelines and compliance measures because meaningless guidelines (as broad goals) result in insignificant outcomes improvement (Nicholas *et al.*, 2010; The Economist Intelligence Unit, 2015).

Key to the implementation of clinical guidelines as well as other policies towards value is to measure outcomes, otherwise results on policies' implementation remain unknown. But defining the outcomes that truly matter to patients is a demanding question.

Outcomes measures have not gone far beyond the common analysis of mortality and safety and even quality measures have not been more than an analysis on the compliance to clinical guidelines (with all the limitations stated above) (Porter e Lee, 2013). A remarkable work is being done by International Consortium for Health Outcomes Measurement (ICHOM) with the objective to identify the outcomes that matter most to patients.

The point now is to understand how to collect these outcomes measures. At the moment outcomes start to be measured, comparisons between institutions are expected to occur as well as changes towards optimization of value (Ackerman e Stowell, 2015; Arora, Hazelzet e Koudstaal, 2016). One point on the collection of outcomes measures is that it should be centered on condition or on the patient and not on the specialty or procedure and it should cover the full cycle of care for the condition and follow patient aftercare (Porter e Lee, 2013).

In fact, healthcare is not expected to re-invent the wheel. There are a lot of providers that already collect several outcomes measures. The question is to re-organize the data that is already collected, considering data that is deemed relevant for outcomes analysis, dropping data that may be irrelevant and starting to collect data that truly matters. Moreover, it is especially relevant to analyze the data that exists but is not available to general research.

The last decade was marked by the financial crisis and the search for efficiencies in all markets with healthcare not being an exception. The emerging paradigm of value estimation and the focus on the patient demands for a change in the structure of healthcare structure as we know it (Porter e Lee, 2013; The Economist Intelligence Unit, 2016a). Three pillars of structural change are identified: incentives on value activities, patient empowerment and re-structuration of healthcare.

While studying variations in medical practice, the importance of incentives has been documented. Increment or reduction of specific activity was possible if the right incentives were implemented. This was true for influencing activity volume but also as identification of the reasoning behind variations. Several authors have argued that the amount of care depends on the resources available. While it was not possible to clearly define which resources influence A-CSR in Portugal, resource availability could explain part of the variations. But a significant part of the variations is expected to be influenced by incentives, especially financial incentives. In the shift towards value and with the collection of outcome measures, it is expected that a new system of incentives, for example on reimbursement schemes, promotes value while reducing overall spending (Ackerman e Stowell, 2015; Dunbar-Rees, Panch e Dancy, 2014). Like before, an integrated perspective of conditions and patient would be preferable to the actual fee-for-service model.

Financial incentives should not be confounded with financial measures though. In the transition for a Value Based Healthcare Model, financial incentives would prompt value over activity, physicians and healthcare in general would prefer to optimize value instead of activity per se. What we have seen in Portugal is a bundle of financial measures focused on lowering costs more than to promote value (The Economist Intelligence Unit, 2016b).

The second point on the transformation of healthcare is the evolution of patient empowerment. One of the reasons that explain variations in medical practice is this agency relation between physicians and patients. Agency relation may be distorted when physicians have incentives to break the agency relation they hold with the patients. When patients are put on the center of care, when the outcomes measured are the outcomes that matter to patients and more important when patient empowerment is developed agency relationship become softer preventing physicians to act on their own interest (Dunbar-Rees *et al.*, 2014).

And at last, the change in the structure of healthcare. Looking at the patient at the center of care and to outcome measures that matter to patients, requires re-organizing care around an integrated view of the patient. Up until now, healthcare institutions have been organized towards specialties but this may no longer be sustainable (Dunbar-Rees *et al.*, 2014). An example of re-organization is the shift from specialties organizations towards disease management organizations with multidisciplinary teams integrating several physicians. Still a lot needs to be studied on this re-structure.

In the last decade, healthcare has been increasingly struggling with pressures to reduce costs especially since the deployment of financial crisis. Discussions about waste, inefficiencies and optimization of resources have promoted research on several areas, namely on variations in medical practice and efficiency estimation. Identification of low-value care and misuse of resources have brought the discussion to a different level: should we be looking to reduce waste and increment efficiency or a new perspective on healthcare is needed? Should we be looking for activity or should we be looking for value of the care provided? While healthcare has seen an impressive development in terms of technologies, healthcare structure has not changed for decades. Maybe it is the time to re-structure healthcare as we know it, putting the patient at the center, looking for outcomes that truly matter to patients and finally pursuing value in healthcare.

7. Conclusions

Variations in medical practice exist in Portugal but the magnitude and evolution vary by procedure in analysis. For catheterization and percutaneous transluminal coronary angioplasty geographical variations reduced but for knee arthroscopy and knee replacement geographic variations are persistently high. In cases where variation rates are still high despite the overall reduction in the number of procedures performed there is a sign that clinical harmonization is required.

When comparing countries, the scope of the procedure (e.g. low-risk C-sections, clinical indication) should also be clearly stated to avoid non-comparable comparisons. In spite of the potential drawbacks and cautions required by international comparisons, these analyses provide a benchmarking that flags opportunities to improve country performances.

Availability of resources affect medical practice in a modest extent and so, policies on resource affection may result in modest results. Policies on medical incentives towards an objective may be more effective.

Stochastic frontier analysis is a good analytical tool to compare hospital's efficiency levels. When applying this technique, it is found that efficiency levels of Portuguese hospitals are not homogenous even though inefficiencies can be attributed to random shocks out of hospital control. Number of beds appears to be a significant driver on inefficiency estimation.

There are several questions that arise from the works developed in this PhD thesis, namely: i) the measures we are currently using to address patient access to care; ii) the sufficiency of standardization methods; iii) the way we are addressing quality of care; iv) the lack of information on true clinical outcomes, in contrast with quantity of care provided; v) relevant data availability.

The methodologies used to identify variations in healthcare are indeed a starting point to perceive heterogeneity in healthcare provision, but the study of variations is medical practice should include not only the definition of an optimal threshold but the definition of the appropriate population to benefit from healthcare. This is essential to understand which geographical areas are performing optimally and the excess consumption in different regions not based on a number but on population needs.

As for the analysis of the quality and efficiency of healthcare systems, the dichotomy "quantity of care produced" and "available resources" should be improved to address topics such as "clinical outcomes" and "value of care". When this happens, we shall leave the concept of how much patients should we be treating to how should we be treating the patients.

This implies reformulating what we are looking for in healthcare provision and which relevant data we collect. When policies are designed towards optimization of clinical outcomes (for example with financial incentives) we shall expect collection of more detailed data than mortality and safety and an optimization of clinical value.

8. References

ACKERMAN, C.; STOWELL, C. - **Measuring Outcomes The Key to Value Based Health Care**

AELVOET, W. *et al.* - Screening for inter-hospital differences in cesarean section rates in low-risk deliveries using administrative data: an initiative to improve the quality of care. **BMC health services research**. ISSN 1472-6963 , *8*, 3. doi: 10.1186/1472-6963-8-3.

AIGNER, D.; LOVELL, C. A. K.; SCHMIDT, P. - Formulation And estimation Of Stochastic Frontier Production Function Models. **Journal of Econometrics**. ISSN 03044076 , *6*(1), 21–37. doi: 10.1016/0304-4076(77)90052-5.

ALARI, A.; LAFORTUNE, G.; SRIVASTAVA, D. - Canada: Geographic variations in health care. Em **Geographic Variations in Health Care: What Do We Know and What Can Be Done To Improve Health System Performance?** [S.l.] : OECD Health Policy Studies, 2014 (pp. 137–169).

ALVAREZ-BARTOLOMÉ, M.; GOGORCENA-AOIZ, A. - Spain: Geographic variations in health care. Em **Geographic Variations in Health Care: What Do We Know and What Can Be Done To Improve Health System Performance?** [S.l.] : OECD Health Policy Studies, 2014 (pp. 343–365).

ANTHONY, D. L. *et al.* - How much do patients' preferences contribute to resource use? **Health Affairs**. ISSN 1544-5208 , *28*(3), 864–873. doi: 10.1377/hlthaff.28.3.864.

APPLEBY, J. *et al.* - Variations in Health Care. The good, the bad and the inexplicable. **Variations in health**. , 1–32.

ARORA, J.; HAZELZET, J.; KOUDSTAAL, M. - A blueprint for measuring health care outcomes. **Harvard Business Review**.

ATLAS DE VARIACIONES EN LA PRÁCTICA MÉDICA - **Apotheke AVPM Núm. 5** [Em linha] Disponível em WWW:URL:http://www.atlasvpm.org/documents/10157/16633/Apotheke_N5_metodo+directo+e+indirecto.pdf.

Atlas de variaciones en la práctica médica en el Sistema Nacional de Salud [The Atlas of Medical Practice Variation in the Spanish National Health System] - [Em linha] Disponível em WWW:URL:http://www.atlasvpm.org/.

BAICKER, K. *et al.* - Who You Are And Where You Live: How Race And Geography Affect The Treatment Of Medicare Beneficiaries. **Health Affairs**. ISSN 0278-2715 , *Suppl Vari*, VAR33–44. doi: 10.1377/hlthaff.var.33.

BAICKER, K.; BUCKLES, K.; CHANDRA, A. - Geographic variation in the appropriate use of cesarean delivery. **Health Affairs**. ISSN 02782715 , *25*(5). doi: 10.1377/hlthaff.25.w355.

BAICKER, K.; CHANDRA, A. - Medicare spending, the physician workforce, and beneficiaries' quality of care. **Health Affairs**. ISSN 02782715 , *23*(SUPPL.). doi: 10.1377/hlthaff.W4.184.

BANKER, R. D.; CHARNES, A; COOPER, W. W. - Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis SOME MODELS FOR ESTIMATING TECHNICAL AND SCALE INEFFICIENCIES IN DATA ENVELOPMENT ANALYSIS. **Management**. ISSN 00251909 , *30*(9), 1078–1092. doi: 10.1287/mnsc.30.9.1078.

BARROS, P. - Economia da saúde: conceitos e comportamentos. Olhares sobre a saúde. . Coimbra : Edições Almedina, 2009. (2ª. ed.)

BERNAL-DELGADO, E. *et al.* - ECHO: Health care performance assessment in several European health systems. **European Journal of Public Health**. ISSN 1464360X , *25*, 3–7. doi: 10.1093/eurpub/cku219.

BERNAL-DELGADO, E.; GARCÍA-ARMESTO, S.; PEIRÓ, S. - Atlas of Variations in Medical Practice in Spain: The Spanish National Health Service under scrutiny. **Health Policy**. ISSN 01688510 , *114*(1), 15–30. doi: 10.1016/j.healthpol.2013.07.013.

BERNAL-DELGADO, E.; MARTÍNEZ-LIZAGA, N.; RIDAO-LÓPEZ, M. - **Análisis del exceso de consume de recursos en pacientes con enfermedad tromboembolica tras cirugía**

BERWICK, D.; HACKBARTH, A. - Eliminating waste in US health care. **JAMA : the journal of the American Medical Association**. ISSN 1538-3598 , *307*(14), 1513–6. doi: 10.1001/jama.2012.362.

BIRKMEYER, J. *et al.* - Understanding of regional variation in the use of surgery. **The Lancet**. ISSN 01406736 , *382*(9898), 1121–1129. doi: 10.1016/S0140-6736(13)61215-5.

BOJKE, C. *et al.* - Regional variation in the productivity of the English National Health Service. **Health Economics**. ISSN 10579230 , *22*(2), 194–211. doi: 10.1002/hec.2794.

BROWN, H. S. - Physician demand for leisure: Implications for cesarean section rates. **Journal of Health Economics**. ISSN 01676296 , *15*(2), 233–242. doi: 10.1016/0167-6296(95)00039-9.

BUSSE, R.; QUENTIN, W. - Moving towards transprency, efficiency and quality in hospitals: Conclusions and recommendations. Em BUSSE, R. *et al.* (Eds.) - **Diagnosis-Related Groups in Europe - moving towards transparency, efficiency and quality in hospitals**. [S.l.] : McGraw Hill, 2011

CAREY, K.; DOR, A. - Expense preference behavior and management «outsourcing»: A comparison of adopters and non-adopters of contract management in U.S. hospitals. **Journal of Productivity Analysis**. ISSN 0895562X , *29*(1), 61–75. doi: 10.1007/s11123-007-0065-3.

CARINCI, F.; STANISLAO, F.; MOIRANO, F. - Italy: Geographic variations in health care. Em **Geographic Variations in Health Care: What Do We Know and What Can Be Done To Improve Health System Performance?** [S.l.] : OECD Health Policy Studies, 2014 (pp. 285–316).

CARLSEN, F.; GRYTTEN, J. - More physicians: Improved availability or induced demand? **Health Economics**. ISSN 10579230 , *7*(6), 495–508. doi: 10.1002/(SICI)1099-1050(199809)7:6<495::AID-HEC368>3.0.CO;2-S.

CHANDRA, A.; CUTLER, D.; SONG, Z. - Who Ordered That? The Economics of Treatment Choices in Medical Care. **Handbook of Health Economics**. ISSN 15740064 , *2*, 397–432. doi: 10.1016/B978-0-444-53592-4.00006-2.

CHARNES, A.; COOPER, W. W.; RHODES, E. - Measuring the Efficiency of Decision Making Units, short communication. **European Journal of Operational Research**. , *3*, 339. doi: http://dx.doi.org/10.1016/0377-2217(78)90138-8.

CHASSIN, M. R.; MCCUE, S. M. - Randomized Trial of Medical Quality Assurance: Improving Physicians' Use of Pelvimetry. **The Journal of American Medical Association**. ISSN 15383598 , *256*(8), 1012–1016. doi: 10.1001/jama.1986.03380080058028.

CLARK, S. L. *et al.* - Institutional influences on the primary cesarean section rate in Utah, 1992 to 1995. **American Journal of Obstetrics and Gynecology**. ISSN 00029378 , *179*(4), 841–845. doi: 10.1016/S0002-9378(98)70175-9.

COELLI, T. J.; PRASADA RAO, D. S.; BATTESE, G. E. - An introduction to efficiency and productivity analysis. . Boston : Kluwer Academic Publishers, 1998.

COLLA, C. H. *et al.* - Choosing Wisely: Prevalence and Correlates of Low-Value Health Care Services in the United States. **Journal of General Internal Medicine**. ISSN 15251497 , *30*(2), 221–228. doi:

10.1007/s11606-014-3070-z.

CONSELHO SUPERIOR DE ESTATÍSTICA - Tipologia de áreas urbanas (TIPAU). **Deliberação da Secção Permanente de Coordenação Estatística**. , (Diário da República: 28 de Setembro de 2009), 2ª série, nº188.

CORALLO, A. N. *et al.* - A systematic review of medical practice variation in OECD countries. **Health policy (Amsterdam, Netherlands)**. ISSN 1872-6054 , *114*(1), 5–14. doi: 10.1016/j.healthpol.2013.08.002.

CROMWELL, J.; MITCHELL, J. B. - Physician-induced demand for surgery. **Journal of Health Economics**. ISSN 01676296 , *5*(4), 293–313. doi: 10.1016/0167-6296(86)90006-8.

CURTIN, L.; KLEIN, R. - Direct Standardization (Age-Adjusted Death Rates). **Healthy People 2000: Statistical Notes**. , *6*, 1–11. doi: 10.1037/e584012012-001.

CYLUS, J.; PAPANICOLAS, I.; SMITH, P. - A framework for thinking about health system efficiency. Em CYLUS, J.; PAPANICOLAS, I.; SMITH, P. (Eds.) - **Health System Efficiency: How to make measurement matter for policy and management.** [S.l.] : European Observatory on Health Systems and Policies, 2016

CYLUS, J.; PEARSON, M. - Cross-national efficiency comparisons of health systems, subsectors and disease areas. Em CYLUS, J.; PAPANICOLAS, I.; SMITH, P. (Eds.) - **Health System Efficiency: How to make measurement matter for policy and management.** [S.l.] : European Observatory on Health Systems and Policies, 2016

DOR, A. - Non-minimum cost functions and the stochastic frontier: On applications to health care providers. **Journal of Health Economics**. ISSN 01676296 , *13*(3), 329–334. doi: 10.1016/0167-6296(94)90032-9.

DRANOVE, D.; WEHNER, P. - Physician-induced demand for childbirths. **Journal of Health Economics**. ISSN 01676296 , *13*(1), 61–73. doi: 10.1016/0167-6296(94)90004-3.

DUNBAR-REES, R.; PANCH, T.; DANCY, M. - From volume to value? can a value-based approach help deliver the ambitious aims of the NHS cardiovascular disease outcomes strategy? **Heart**. ISSN 1355-6037 , *100*(11), 827–832. doi: 10.1136/heartjnl-2013-305269.

DYCK, F. J. *et al.* - Effect of surveillance on the number of hysterectomies in the province of Saskatchewan. **New England Journal of Medicine**. ISSN 0028-4793 , *296*(23), 1326–1328. doi: 10.1056/NEJM197706092962306.

ECHO - **European Collaboration for Healthcare Optimization Handbook on methodology: ECHO performance model; 2014 Apr 27** [Em linha] [Consult. 7 dez. 2016]. Disponível em WWW:URL:http://www.echo-health.eu/handbook/performance_model.html.

ERSKINE, A. *et al.* - How an Early Adopter of Electronic Health Records Uses Big Data. **Harvard Business Review Digital Articles**. , 2–5.

ESCARCE, J. J. - Explaining the association between surgeon supply and utilization. **Inquiry**. ISSN 00469580 , *29*(4), 403–415.

European Collaboration for Health Optimization (ECHO) project - [Em linha] Disponível em WWW:URL:http://echo-health.eu.

European Health Care Optimization, Performance and Efficiency (EuroHOPE) - [Em linha] Disponível em WWW:URL:http://www.eurohope.info/.

EVANS, R. G. - Supplier-Induced Demand: Some Empirical Evidence and Implications. Em **The Economics of Health and Medical Care** [Em linha]. London : Palgrave Macmillan UK, 1974 Disponível em WWW:URL:http://link.springer.com/10.1007/978-1-349-63660-0_10. (pp. 162–173).

FAREBROTHER, J. - United Kingdom (England): Geographic variations in health care. Em OECD HEALTH POLICY STUDIES (Ed.) - **Geographic Variations in Health Care: What Do We Know and What Can Be Done To Improve Health System Performance?** (pp. 393–415).

FARRELL, M. J. - The Measurement of Productive Efficiency. **Journal of the Royal Statistical Society. Series A (General)**. ISSN 00359238 , *120*(3), pp. 253–290. doi: 10.1016/S0377-2217(01)00022-4.

FARSI, M. - The temporal variation of cost-efficiency in Switzerland's hospitals: An application of mixed models. **Journal of Productivity Analysis**. ISSN 0895562X , *30*(2), 155–168. doi: 10.1007/s11123-008-0105-7.

FARSI, M.; FILIPPINI, M. - **An analysis of efficiency and productivity in Swiss hospitals: Quaderno N. 05-01** [Em linha] Disponível em WWW:URL:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.429.8586&rep=rep1&type=pdf.

FELDMAN, R.; SLOAN, F. - Competition among physicians, revisited. **Journal of health politics, policy and law**. ISSN 0361-6878 , *13*(2), 239–261. doi: 10.1215/03616878-13-2-239.

FISHER, E. *et al.* - The implications of regional variations in Medicare spending. **Annals of Internal Medicine**. ISSN 1539-3704 , *138*, 273–311. doi: 10.7326/0003-4819-138-4-200302180-00007.

FOLLAND, S.; GOODMAN, A.; STANO, M. - The Economics of Health and Health Care 7th Edition. [Em linha]. [S.l.] : Pearson, 2013. (7th. ed.) Disponível em WWW:URL:ftp://174.46.176.136/Pearson/9780132773690.pdf.ISBN 978-0-13-277369-0.

FOLLAND, S.; HOFLER, R. - How reliable are hospital efficiency estimates? exploiting the dual to homothetic production. **Health Economics**. ISSN 10579230 , *10*(8), 683–698. doi: 10.1002/hec.600.

FOLLAND, S.; STANO, M. - Small area variations: a critical review of propositions, methods, and evidence. **Medical care review**. ISSN 0025-7087 , *47*(4), 419–65.

FRANCO, F.; FORTUNA, M. - **O método de fronteira estocástica na medição da eficiência dos serviços hospitalares: uma revisão bibliográfica. Documento de trabalho nº2/2003** [Em linha] Disponível em WWW:URL:http://apes.pt/wp-content/uploads/2015/05/dt_022003.pdf.

FUCHS, V. R. - The supply of surgeons and the demand for operations. **The Journal of human resources**. ISSN 0022-166X , *13 Suppl*, 35–56. doi: 10.2307/145247.

GANNON, B. - Testing for variation in technical efficiency of hospitals in Ireland. **Economic and Social Review**. ISSN 00129984 , *36*(3), 273–294.

GARCÍA-ALTÉS, A. *et al.* - Measuring the performance of health care services: a review of international experiences and their application to urban contexts. **Gaceta sanitaria / S.E.S.P.A.S**. ISSN 0213-9111 , *20*(4), 316–24. doi: 10.1157/13091148.

GLOVER, J. A. - The incidence of tonsillectomy in school children. **The Indian Journal of Pediatrics**. ISSN 00195456 , *5*(4), 252–258. doi: 10.1007/BF02751831.

GONÇALVES, R.; BARROS, P. - Economies of scale and scope in the provision of diagnostic techniques and therapeutic services in Portuguese hospitals. **Applied Economics**. ISSN 0003-6846 , *45*(4), 415–433. doi: 10.1080/00036846.2011.605758.

GREEN, L. A.; BECKER, M. P. - Physician decision making and variation in hospital admission rates for suspected acute cardiac ischemia. A tale of two towns. **Medical care**. ISSN 0025-7079 (Print) , *32*(11), 1086–1097.

GROENEWEGEN, P. P.; WESTERT, G. P. - Is there a time trend in medical practice variations?: A review of the literature and a critical analysis of theoretical approaches . **J Public Health**. , *12*(3), 229–236.

GRUBER, J.; OWINGS, M. - Physician financial incentives and cesarean section delivery. **The Rand journal of economics**. ISSN 0741-6261 , *27*(1), 99–123. doi: 10.2307/2555794.

GRYTTEN, J.; CARLSEN, F.; SØRENSEN, R. - Supplier inducement in a public health care system. **Journal of Health Economics**. ISSN 01676296 , *14*(2), 207–229. doi: 10.1016/0167-6296(94)00036-4.

GRYTTEN, J.; MONKERUD, L.; SORENSEN, R. - Adoption of Diagnostic Technology and Variation in Caesarean Section Rates: A Test of the Practice Style Hypothesis in Norway. **Health Services Research**. ISSN 00179124 , *47*(6), 2169–2189. doi: 10.1111/j.1475-6773.2012.01419.x.

GUSMANO, M. K. *et al.* - Disparities in access to health care in three French regions. **Health Policy**. ISSN 01688510 , *114*(1), 31–40. doi: 10.1016/j.healthpol.2013.07.011.

HADLEY, J.; ZUCKERMAN, S. - The role of efficiency measurement in hospital rate setting. **Journal of Health Economics**. ISSN 01676296 doi: 10.1016/0167-6296(94)90033-7.

HART, O.; HOLMSTROM, B. - A Theory of Firm Scope. **Source: The Quarterly Journal of Economics**. ISSN 00335533, 15314650 , *125*(2), 483–513. doi: 10.1162/qjec.2010.125.2.483.

HERR, A. - Cost and technical efficiency of German hospitals: Does ownership matter? Em **Health Economics**ISBN 1099-1050; 1057-9230 (Vol. 17, pp. 1057–1071).

HERWARTZ, H.; STRUMANN, C. - On the effect of prospective payment on local hospital competition in Germany. **Health Care Management Science**. ISSN 13869620 , *15*(1), 48–62. doi: 10.1007/s10729-011-9180-9.

HINDLE, D. - Implementing DRGs in Slovenia: why the Australian variant was selected. **Australian Health Review**. , *26*(3).

HO, K.; PAKES, A. - Do physician incentives affect hospital choice? A progress report. Em **International Journal of Industrial Organization**ISBN 0167-7187 (Vol. 29, pp. 317–322).

HOLLINGSWORTH, B. - Health system efficiency: measurement and policy. Em CYLUS, J.; PAPANICOLAS, I.; SMITH, P. (Eds.) - **Health System Efficiency: How to make measurement matter for policy and management.** [S.l.] : European Observatory on Health Systems and Policies, 2016

HOLLINGSWORTH, B.; DAWSON, P. J.; MANIADAKIS, N. - Efficiency measurement of health care: a review of non-parametric methods and applications. **Health care management science**. ISSN 1386-9620 , *2*(3), 161–172. doi: 10.1023/A:1019087828488.

IOM - Institute of Medicine: Target Decision-making, Not Geography. Em NEWHOUSE, J. P.; GARBER, R.; GRAHAM (Eds.) - **Institute of Medicine Final Report, Committee on Geographic Variation in Health Care Spending and Promotion of High-Value Care**. Board on H ed. (Board on H ed.).

IVERSEN, T. - The effects of a patient shortage on general practitioners' future income and list of patients. **Journal of Health Economics**. ISSN 01676296 doi: 10.1016/j.jhealeco.2003.10.001.

JACKSON, T. *et al.* - Measurement of adverse events using «incidence flagged» diagnosis codes. **Journal of health services research & policy**. ISSN 1355-8196 , *11*(1), 21–6. doi: 10.1258/135581906775094271.

JACKSON, T. *et al.* - Marginal costs of hospital-acquired conditions: information for priority-setting for patient safety programmes and research. **Journal of health services research & policy**. ISSN 1758-1060 , *16*(3), 141–6. doi: 10.1258/jhsrp.2010.010050.

JACOBS, R.; SMITH, P. C.; STREET, A. - Measuring efficiency in health care: Analytic techniques and health policy. Measuring Efficiency in Health Care: Analytic Techniques and Health Policy [Em linha] Disponível em WWW:URL:http://www.scopus.com/inward/record.url?eid=2-s2.0-84926079120&partnerID=40&md5=17fd68a11b342c8c6103961c76471cd8.ISBN 9780511617492 (ISBN);

0521851440 (ISBN); 9780521851442 (ISBN).

JOHNSON, E. M. - Physician-Induced Demand. **Encyclopedia of Health Economics**. ISSN 01676296 , *1*, 77–82. doi: 10.1016/0167-6296(82)90002-9.

KANG, H.-C.; HONG, J.-S.; PARK, H.-J. - Development of Peer-Group-Classification Criteria for the Comparison of Cost Efficiency among General Hospitals under the Korean NHI Program. **Health Services Research**. ISSN 00179124 , *47*(4), 1719–1738. doi: 10.1111/j.1475-6773.2012.01379.x.

KITTELSEN, S. *et al.* - Hospital productivity and the Norwegian ownership reform – A Nordic comparative study. **HERO On line Working Paper Series**.

KITTELSEN, S. *et al.* - Decomposing the productivity differences between hospitals in the Nordic countries. **Journal of Productivity Analysis**. ISSN 0895-562X, 1573-0441 , *43*(3), 281–293. doi: 10.1007/s11123-015-0437-z.

KOBEL, C. *et al.* - DRG systems and similar patient classification systems in Europe. Em BUSSE, R. *et al.* (Eds.) - **Diagnosis-Related Groups in Europe - moving towards transparency, efficiency and quality in hospitals** [Em linha]. [S.l.] : McGraw Hill, 2011 Disponível em WWW:URL:http://www.euro.who.int/__data/assets/pdf_file/0004/162265/e96538.pdf. (pp. 37–58).

KOOREMAN, P. - Nursing home care in the Netherlands: A nonparametric efficiency analysis. **Journal of Health Economics**. ISSN 01676296 , *13*(3), 301–316. doi: doi:10.1016/0167-6296(94)90029-9.

KRISTENSEN, T. *et al.* - Economies of scale and scope in the Danish hospital sector prior to radical restructuring plans. **Health Policy**. ISSN 01688510 , *106*(2), 120–126. doi: 10.1016/j.healthpol.2012.04.001.

KUMAR, S. - Specialty hospitals emulating focused factories. **International Journal of Health Care Quality Assurance**. ISSN 0952-6862 , *23*(1), 94–109. doi: 10.1108/09526861011010703.

LABELLE, R.; STODDART, G.; RICE, T. - A re-examination of the meaning and importance of supplier-induced demand. **Journal of Health Economics**. ISSN 01676296 , *13*(3), 347–368. doi: 10.1016/0167-6296(94)90036-1.

LAUER, J. A *et al.* - Determinants of caesarean section rates in developed countries : supply , demand and opportunities for control. **World Health Organization**. , 22.

LI, T.; ROSENMAN, R. - Cost inefficiency in Washington hospitals: a stochastic frontier approach using panel data. **Health Care Management Science**. ISSN 1386-9620 , *4*(2), 73–81.

LINNA, M. *et al.* - Measuring cost efficiency in the Nordic Hospitals-a cross-sectional comparison of public hospitals in 2002. **Health Care Management Science**. ISSN 13869620 , *13*(4), 346–357. doi: 10.1007/s10729-010-9134-7.

LINNA, M.; HAKKINEN, U.; MAGNUSSEN, J. - Comparing hospital cost efficiency between Norway and Finland. **Health Policy**. ISSN 01688510 , *77*(3), 268–278. doi: 10.1016/j.healthpol.2005.07.019.

LUDWIG, M.; GROOT, W.; MERODE, F. VAN - Hospital efficiency and transaction costs: A stochastic frontier approach. **Social Science & Medicine**. ISSN 02779536 , *69*(1), 61–67. doi: 10.1016/j.socscimed.2009.04.026.

LUDWIG, M.; MERODE, F. VAN; GROOT, W. - Principal agent relationships and the efficiency of hospitals. **European Journal of Health Economics**. ISSN 16187598 , *11*(3), 291–304. doi: 10.1007/s10198-009-0176-z.

MATEUS, C. *et al.* - Portugal: Geographic variations in health care. Em **Geographic Variations in Health Care: What Do We Know and What Can Be Done To Improve Health System Performance?** [S.l.] : OECD Health Policy Studies, 2014 (pp. 317–342).

MCGUIRE, A.; HENDERSON, J.; MOONEY, C. - The Economics of Health Care: an Introductory Text. . London : Routledge and Keegan, 1988.

MCGUIRE, T. - Chapter 9 Physician agency. **Handbook of Health Economics**. ISSN 15740064 doi: 10.1016/S1574-0064(00)80168-7.

MCGUIRE, T.; PAULY, M. - Physician response to fee changes with multiple payers. **Journal of Health Economics**. ISSN 01676296 , *10*(4), 385–410. doi: 10.1016/0167-6296(91)90022-F.

MCPHERSON, K. *et al.* - Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway. **New England Journal of Medicine**. ISSN 0028-4793 , *307*(21), 1310–4. doi: 10.1056/NEJM198211183072104.

MEDIN, E. *et al.* - Cost efficiency of university hospitals in the Nordic countries: a cross-country analysis. **The European journal of health economics : HEPAC : health economics in prevention and care**. ISSN 1618-7601 , *12*(6), 509–19. doi: 10.1007/s10198-010-0263-1.

MEDIN, E. *et al.* - International hospital productivity comparison: Experiences from the Nordic countries. **Health Policy**. ISSN 01688510 , *112*(1-2), 80–87. doi: 10.1016/j.healthpol.2013.02.004.

Medizinischen Versorgungatlas in Deutschland [Atlas of Medical Care in Germany] - [Em linha] Disponível em WWW:URL:http://www.versorgungsatlas.de/.

MEEUSEN, W.; BROECK, J. VAN DEN - Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. **International Economic Review**. ISSN 00206598 , *18*(2), 435. doi: 10.2307/2525757.

MORIATES, C.; ARORA, V.; SHAH, N. - Understanding Value-Based Healthcare. . [S.l.] : Mc Graw Hill, 2015.

NASSIRI, A.; ROCHAIX, L. - Revisiting physicians' financial incentives in Quebec: A panel system approach. **Health Economics**. ISSN 10579230 , *15*(1), 49–64. doi: 10.1002/hec.1012.

NEWHOUSE, J. P. - Medical care costs: how much welfare loss? **The Journal of Economic Perspectives**. ISSN 0895-3309 doi: 10.1257/jep.6.3.3.

NEWHOUSE, J. P. - Frontier estimation: How useful a tool for health economics? **Journal of Health Economics**. ISSN 01676296 , *13*(3), 317–322. doi: 10.1016/0167-6296(94)90030-2.

NGUYEN, N. X.; DERRICK, F. W. - Physician behavioral response to a Medicare price reduction. **Health services research**. ISSN 0017-9124 , *32*(3), 283–298.

NHS Atlas of Variations in Healthcare - [Em linha] Disponível em WWW:URL:https://www.england.nhs.uk/rightcare/intel/cfv/atlas/.

NICHOLAS, L. H. *et al.* - Hospital process compliance and surgical outcomes in medicare beneficiaries. **Archives of surgery (Chicago, Ill. : 1960)**. ISSN 1538-3644 , *145*(10), 999–1004. doi: 10.1001/archsurg.2010.191.

OECD - **Geographic Variations in Health Care: What Do We Know and What Can Be Done to Improve Health System Performance?**

OLIVEIRA, M. D.; BEVAN, G. - Modelling hospital costs to produce evidence for policies that promote equity and efficiency. **European Journal of Operational Research**. ISSN 03772217 , *185*(3), 933–947. doi: 10.1016/j.ejor.2006.02.053.

OR, Z. *et al.* - Diagnosis related groups and variations in resource use for child delivery across 10 European countries. **Health Economics (United Kingdom)**. ISSN 10579230 , *21*(SUPPL. 2), 55–65. doi: 10.1002/hec.2835.

OR, Z.; VERBOUX, D. - France: Geographical variations in health care. Em **Geographic Variations in Health Care: What Do We Know and What Can Be Done To Improve Health System Performance?** [S.l.] : OECD Health Policy Studies, 2014 (pp. 221–244).

OXLEY, H.; MACFARLEN, M. - **Health Care Reform Controlling Spending and Increasing Efficiency**. OECD Publishing, Paris : [s.n.]

PAN-AMERICAN HEALTH ORGANIZATION - Standardization: a classic epidemiological method for the comparison of rates. **Epidemiological bulletin**. ISSN 0256-1859 , *23*(3), 9–12.

PAPANICOLAS, I.; SMITH, P. C. - Health System Performance Comparison: An Agenda for Policy, Information and Research. European Observatory on Health Systems and Policies [Em linha] Disponível em WWW:URL:http://www.popline.org/node/535027.ISBN 9780335247264.

PAULY, M. - Doctors and their workshops: Economic models of physician behavior. . London : The University of Chicago Press, 1980. ISBN 9780226650463.

PEIRÓ, S.; MAYNARD, A. - Variations in health care delivery within the European Union. **Eur J Public Health**. ISSN 1101-1262 , *25*(Supp1), 1–2. doi: 10.1093/eurpub/cku223.

PELLEGRINI, S.; KOHLER, D. - Switzerland: Geographic variations in health care. Em **Geographic Variations in Health Care: What Do We Know and What Can Be Done To Improve Health System Performance?** [S.l.] : OECD Health Policy Studies, 2014 (pp. 367–392).

PHELPS, C.; MOONEY, C. - Variations in medical practice use: causes and consequences. Em ARNOULD, R. J.; RICH, R. F.; WHITE, W. (Eds.) - **Competitive Approaches to Health Care Reform**. The Urban ed. Washington, DC : [s.n.] (The Urban ed.).

PLOTZKE, M. R.; COURTEMANCHE, C. - Does procedure profitability impact whether an outpatient surgery is performed at an ambulatory surgery center or hospital? **Health Economics**. ISSN 10579230 , *20*(7), 817–830. doi: 10.1002/hec.1646.

PORTER, M. E.; LEE, T. H. - The Strategy That Will Fix Healthcare. **Harvard Business Review**. ISSN 0017-8012 , *1277*(October), 1–19. doi: 10.1017/CBO9781107415324.004.

PORTER, M. E.; TEISBERG, E. O. - HBR - Michael Porter - Redefining Competition in Healthcare - 2004.pdf. **Harvard Business Review**. ISSN 00178012

QUAST, T.; SAPPINGTON, D. E. M.; SHENKMAN, E. - Does the quality of care in medicaid MCOs vary with the form of physician compensation? **Health Economics**. ISSN 10579230 , *17*(4), 545–550. doi: 10.1002/hec.1264.

RALSTON, S. L. S. *et al.* - Hospital Variation in Health Care Utilization by Children With Medical Complexity. **Pediatrics**. ISSN 1098-4275 , *136*(5), 860–7. doi: 10.1542/peds.2014-3920.

REDISCH, M.; GABEL, J.; BLAXALL, M. - Physician pricing, costs, and income. **Advances in health economics and health services research**. ISSN 0731-2199 , *2*, 197–228.

REINHARDT, U. E. - The theory of physician-induced demand: reflections after a decade. **Journal of health economics**. ISSN 0167-6296 , *4*(2), 187–93.

RETZLAFF-ROBERTS, D.; CHANG, C. F.; RUBIN, R. M. - Technical efficiency in the use of health care resources: a comparison of OECD countries. **Health Policy**. ISSN 01688510 , *69*(1), 55–72. doi: 10.1016/j.healthpol.2003.12.002.

RICE, T. H. - The impact of changing medicare reimbursement rates on physician-induced demand. **Medical care**. ISSN 0025-7079 , *21*(8), 803–815. doi: 10.1097/00005650-198308000-00004.

RIDAO-LÓPEZ, M. *et al.* - Evaluación de la variabilidad geográfica de la práctica médica como herramienta de evaluación de políticas sanitarias. **Presupuesto y gasto Público**. , *68*, 193–210.

RIZZO, J. A.; BLUMENTHAL, D. - Is the Target Income Hypothesis an Economic Heresy. **Medical Care Research and Review**. ISSN 1077-5587 , *53*(3), 243–266. doi: 10.1177/107755879605300301.

RIZZO, J. A.; ZECKHAUSER, R. J. - Reference Incomes, Loss Aversion, and Physician Behavior. **The Review of Economics and Statistics**. ISSN 0034-6535 , *85*(4), 909–922. doi: 10.1162/003465303772815817.

ROSKO, M. D. - Cost efficiency of US hospitals: a stochastic frontier approach. **Health Economics**. ISSN 1057-9230 , *10*(August 1999), 539–551. doi: 10.1002/hec.607.

ROSKO, M. D.; MUTTER, R. L. - Stochastic frontier analysis of hospital inefficiency: a review of empirical issues and an assessment of robustness. **Medical Care Research & Review**. ISSN 1077-5587 , *65*(2), 131–166.

ROSKO, M. D.; MUTTER, R. L. - What have we learned from the application of stochastic frontier analysis to U.S. hospitals?. **Medical Care Research & Review**. ISSN 1552-6801 , *68*(1 Suppl), 75S–100S.

ROSSITER, L.; WILENSKY, G. - Identification of physician-induced demand. **Journal of Human Resources**. ISSN 0022166X , *19*(2), 231–244. doi: 10.2307/145565.

SCHANG, L. *et al.* - From data to decisions? Exploring how healthcare payers respond to the NHS Atlas of Variation in Healthcare in England. **Health Policy**. ISSN 01688510 , *114*(1), 79–87. doi: 10.1016/j.healthpol.2013.04.014.

SCHLEINIGER, R. - Regional Quantity, Productivity and Efficiency Measures of the Swiss Health Care System. **Swiss Journal of Economics and Statistics (SJES)**. ISSN 03039692 , *144*(III), 459–476. doi: http://www.sjes.ch/index.php.

SHWARTZ, M. *et al.* - Small area variations in hospitalization rates: how much you see depends on how you look. **Medical care**. ISSN 0025-7079 , *32*(3), 189–201. doi: 10.1097/00005650-199403000-00001.

SIROVICH, B. E. *et al.* - Regional variations in health care intensity and physician perceptions of quality of care. **Annals of Internal Medicine**. ISSN 00034819 , *144*(9), 641–649. doi: 10.7326/0003-4819-144-9-200605020-00007.

SKINNER, J. - What do stochastic frontier cost functions tell us about inefficiency? **Journal of Health Economics**. ISSN 01676296 , *13*(3), 323–328. doi: 10.1016/0167-6296(94)90031-0.

SKINNER, J.; WENNBERG, J. - How Much is Enough? Efficiency and Medicare Spending in the Last Six Months of Life. **National Bureau of Economic Research Working Paper Series**. , *6513*. doi: 10.3386/w6513.

SMITH, P. C.; MOSSIALOS, E.; PAPANICOLAS, I. - Performance Measurement for Health System Improvement Experiences, Challenges and Prospects. **World Health**. ISSN 09540962 , *57*(2), 726. doi: 10.1111/j.1467-9302.2005.00476.x.

SNYDER, C. C. *et al.* - The influence of hospital type on induction of labor and mode of delivery. **American Journal of Obstetrics and Gynecology**. ISSN 00029378 , *205*(4). doi: 10.1016/j.ajog.2011.05.004.

SONG, Y. *et al.* - Regional variations in diagnostic practices. **The New England journal of medicine**. ISSN 1533-4406 , *363*(1), 45–53. doi: 10.1056/NEJMsa0910881.

SRIVASTAVA, D. *et al.* - Geographic variations in health care use in 13 countries: A synthesis of findings. Em **Geographic Variations in Health Care: What Do We Know and What Can Be Done To Improve Health System Performance?** [S.l.] : OECD Health Policy Studies, 2014ISBN 9789264216587 (pp. 27–77).

STANO, M. - An analysis of the evidence on competition in the physician services markets. **Journal of Health Economics**. ISSN 01676296 , *4*(3), 197–211. doi: 10.1016/0167-6296(85)90029-3.

STREET, A. *et al.* - DRG-based hospital payment and efficiency: theory, evidence, and challenges. Em **Diagnosis-Related Groups in Europe - moving towards transparency, efficiency and quality in hospitals**. [S.l.] : European Observatory on Health Systems and Policies, 2011

SUND, R.; HÄKKINEN, U. - Unsing registry data to compare health care efficiency. Em CYLUS, J.; PAPANICOLAS, I.; SMITH, P. (Eds.) - **Health System Efficiency: How to make measurement matter for policy and management.** [S.l.] : European Observatory on Health Systems and Policies, 2016

SUNDMACHER, L.; BUSSE, R. - Geographic variation in health care—A special issue on the 40th anniversary of «Small area variation in health care delivery». **Health Policy**. ISSN 01688510 , *114*(1), 3–4. doi: 10.1016/j.healthpol.2013.11.011.

The Dartmouth Atlas of Health Care - [Em linha] Disponível em WWW:URL:http://www.dartmouthatlas.org/.

THE ECONOMIST INTELLIGENCE UNIT - An introduction to value-based healthcare in Europe. **The Economist**. , *May*, 1–6. doi: 999/UK/16-02/NMAR/1003c.

THE ECONOMIST INTELLIGENCE UNIT - **Value-based healthcare in Europe - Laying the foudations**

THE ECONOMIST INTELLIGENCE UNIT - **Value-Based Healthcare in Portugal - Necessity is the mother of invention**

VITALIANO, D. F.; TOREN, M. - Cost and efficiency in nursing homes: a stochastic frontier approach. **Journal of health economics**. ISSN 0167-6296 , *13*, 281–300; discussion 341–343. doi: 10.1016/0167-6296(94)90028-0.

VOLINN, E. *et al.* - Why does geographic variation in health care practices matter? (And seven questions to ask in evaluating studies on geographic variation). **Spine**. ISSN 03622436

VOORDE, C. De VAN; DOORSLAER, E. VAN; SCHOKKAERT, E. - Effects of cost sharing on physician utilization under favourable conditions for supplier-induced demand. **Health Economics**. ISSN 10579230 , *10*(5), 457–471. doi: 10.1002/hec.631.

WAGSTAFF, A.; LÓPEZ, G. - Hospital costs in Catalonia: a stochastic frontier analysis. **Applied Economics Letters**. ISSN 1350-4851 , *3*(7), 471–474. doi: 10.1080/758540809.

WANG, J.; ZHAO, Z.; MAHMOOD, A. - **Relative Efficiency, Scale Effect, and Scope Effect of Public Hospitals: Evidence from Australia. Discussion Paper No. 2520** [Em linha] Disponível em WWW:URL:https://papers.ssrn.com/sol3/papers.cfm?abstract_id=956386.

WELCH, H. *et al.* - Geographic Variation in Diagnosis Frequency and Risk of Death Among Medicare Beneficiaries. **JAMA: The Journal of the American Medical Association**. ISSN 1538-3598 , *305*(11), 1113–1118. doi: 10.1001/jama.2011.307.

WENNBERG, J. - Dealing with medical practice variations: A proposal for action. **Health Affairs**. ISSN 02782715 , *3*(2), 6–32. doi: 10.1377/hlthaff.3.2.6.

WENNBERG, J. - Unwarranted variations in healthcare delivery: implications for academic medical centres. **British Medical Journal**. ISSN 09598138 , *325*(April), 961–964. doi: 10.1136/bmj.325.7370.961.

WENNBERG, J. - Forty years of unwarranted variation—And still counting. **Health Policy**. ISSN 01688510 , *114*(1), 1–2. doi: 10.1016/j.healthpol.2013.11.010.

WENNBERG, J.; BARNES, B. A.; ZUBKOFF, M. - Professional uncertainty and the problem of supplier-induced demand. **Social Science and Medicine**. ISSN 02779536 , *16*(7), 811–824. doi: 10.1016/0277-9536(82)90234-9.

WENNBERG, J.; COOPER, M. - The Dartmouth Atlas of Health Care. . [S.l.] : American Hospital Publishing, Inc, 1996.

WENNBERG, J.; FISHER, E.; SKINNER, J. S. - Geography and the debate over medicare reform. **Health Affairs**. ISSN 02782715 doi: 10.1017/CBO9781107415324.004.

WENNBERG, J.; FOWLER, F. J. - A test of consumer contribution to small area variations in health care delivery. **The Journal of the Maine Medical Association**. ISSN 0025-0694 , *68*(8), 275–9.

WENNBERG, J.; FREEMAN, J.; CULP, W. - ARE HOSPITAL SERVICES RATIONED IN NEW HAVEN OR OVER-UTILISED IN BOSTON? **The Lancet**. ISSN 01406736 , *329*(8543), 1185–1189. doi: 10.1016/S0140-6736(87)92152-0.

WENNBERG, J.; GITTELSOHN, A. - Small area variations in health care delivery. **Science (New York, N.Y.)**. ISSN 0036-8075 , *182*(4117), 1102–8. doi: 10.1126/science.182.4117.1102.

WENNBERG, J.; GITTELSOHN, A. - Variations in medical care among small areas. **Scientific American**. ISSN 0036-8733 , *246*(4), 120–134. doi: 10.1038/scientificamerican0482-120.

YASAITIS, L. C.; BYNUM, J. P. W.; SKINNER, J. S. - Association between physician supply, local practice norms, and outpatient visit rates. **Medical care**. ISSN 1537-1948 , *51*(6), 524–31. doi: 10.1097/MLR.0b013e3182928f67.

YIP, W. C. - Physician response to Medicare fee reductions: Changes in the volume of coronary artery bypass graft (CABG)surgeries in the Medicare and private sectors. **Journal of Health Economics**. ISSN 01676296 doi: 10.1016/S0167-6296(98)00024-1.

ZUCKERMAN, S. *et al.* - Clarifying sources of geographic differences in Medicare spending. **The New England journal of medicine**. ISSN 0028-4793 , *363*(1), 54–62. doi: 10.1056/NEJMsa0909253.

ZUCKERMAN, S.; HADLEY, J.; IEZZONI, L. - Measuring hospital efficiency with frontier cost functions. **Journal of Health Economics**. ISSN 01676296 , *13*(3), 255–280. doi: 10.1016/0167-6296(94)90027-2.